CS145 Final Review

Peng Qi, Sophia Nguyen, Karthik Jagadeesh

August 13, 2014

Outline

- Database Systems
- Transactions, Views, and Authorization
- NoSQL

Database Systems *Intro*

- Indexes
- Query Processing and Optimization
- Materialized View
- OLAP (Online Analytical Processing)
- DBMS with Hardware and File Systems, Indexing, Query Processing (in detail), Robustness, Concurrency Control, Transaction Processing, Distributed DBMS, ... $\Rightarrow$ CS245, CS246, CS346

Indexes

Creating Indexes in SQL

```
create index IndexName on T(A1, ..., An)
create unique index IndexName on T(A1, ..., An)
drop index IndexName
```

- Tree-based indexes (Equality and range queries) $O(\log n)$
- Hash-based indexes (Equality queries only) $O(1)$
- Without indexes (Any queries) $O(n)$

Indexes

How can you make life easier for EBayArea/SUPostDoc/CrackerList with indexes? T for tree-based, H for hash-based

- Find item name and description by ItemID ⇒ Item.ItemID (H,T)
- Find buyers' comments by ItemID ⇒ Comments.ItemID (H,T)
- Find inexpensive items (with price range) ⇒ Item.CurrentPrice (T)
- Find items in some category that are not biddable ⇒ Item.CategoryID (H,T), (Item.Biddable (H))
- Find items by name ⇒ Text Indexes, usually as fast as hash-/tree-indexes, but takes up much more space (CS276)

Query Processing

"Go to Meyer Library for office hours at 7"

- AttendOH(Location=MeyerLibrary, Time=7pm) $\Rightarrow$ Query Parsing
- Thorton $\rightarrow$ Seattle | Berkeley | University Ave. | Tressider $\rightarrow$ Meyer Library $\Rightarrow$ Logical Planning & Evaluation
- Walk, Bike, Drive, Take a spaceship... $\Rightarrow$ Physical Planning & Evaluation
- Arrive at Meyer for the OH $\Rightarrow$ Query Execution & Return query results

Query Optimization

Back to our EBayArea problem: Say we are given a user (ID: 123), we want all bid Prices made by this user on an item with SellPrice below $25 where he gave a comment rating less than 3.

Query Optimization

```
select Bid.Price
from Bid, Item, Comments
where Bid.UserID=123 and Item.ItemID=Bid.ItemID
   and Item.SellPrice < 25
   and Comment.Rating < 3
   and Comment.ItemID=Bid.ItemID
   and Comment.UserID=123
```

$$\pi_{Bid.Price}(\sigma_{Bid.UserID=123 \wedge Item.SellPrice<25 \wedge Comment.Rating<3 \wedge Item.ItemID=Bid.ItemID \wedge Comment.UserID=123 \wedge Comment.ItemID=Bid.ItemID}(Bid \times Item \times Price))$$

$$\pi_{Bid.Price}(\sigma_{Bid.UserID=123 \wedge Item.SellPrice<25 \wedge Comment.Rating<3 \wedge Comment.UserID = 123}(Bid \bowtie Item \bowtie Price))$$

$$\pi_{Bid.Price}(\sigma_{Bid.UserID=123} Bid \bowtie \sigma_{Item.SellPrice<25} Item \bowtie \sigma_{Comment.Rating<3 \wedge Comment.UserID=123} Price)$$

Query Optimization

Logical Query Plan

$$\pi_{Bid.Price}(\sigma_{Bid.UserID=123} Bid \bowtie \sigma_{Item.SellPrice<25} Item \bowtie \sigma_{Comment.Rating<3} Price)$$

Physical Query Plan

- Recap for Merge-(Sort-)Join (In class discussion covering Problems 3 & 4 of Problem Set 3)
- What about with indexes?

Materialized Views

```
create materialized view ViewName as
select SomeColumns
from SomeTable
where SomeCondition
```

- What are materialized views good for?
  Frequently computed aggregates, frequently performed joins, (perhaps) frequently needed values that are not in the normal form...
- What should we pay attention to when using materialized views?
  Make sure that insertion, deletion, and update happen correctly in the underlying tables with `INSTEAD OF` triggers.

OLAP[1]

- CUBE (all-you-can-think-of)
- ROLLUP (all-that-makes-sense-in-a-tree)

[1] http://en.wikipedia.org/wiki/Online_analytical_processing

OLAP example

Problem 2a from PSet 3

```
create materialized view T as
select Continent, Country, City, District,
    sum(population) as s
from PopRecord
group by Continent, Country, City, District
    with cube
```

(In class discussion)

OLAP example 2

A slightly modified problem

```
create materialized view T as
select Continent, Country, City, District,
    sum(population) as s
from PopRecord
group by Continent, Country, City, District
    with rollup
```

(In class discussion)