Introduction to Data Management

Lecture #18
(SQL, the Final Chapter…)

Instructor: Mike Carey
mjcarey@ics.uci.edu

Announcements

- Grading progress has been made…!
  - HWs 2 & 3 are now graded
  - Midterm #1 is also graded (!)
- HW #6 coming out tonight, due next Monday
  - The usual one late day will be available as an option
  - The solution will be posted on Tuesday after 5PM
- Midterm #2 will be Wednesday (5/22) in class
  - Relational languages (see syllabus)
  - A sample from last year is available online
- So – let’s finish up relational languages!
Layers of Schemas: Brief “Re-View”

- Many views of one conceptual (logical) schema and an underlying physical schema
  - Views describe how different users see the data.
  - Conceptual schema defines the logical structure of the database
  - Physical schema describes the files and indexes used under the covers
**A Simple View Example (MySQL)**

```sql
CREATE VIEW YoungSailorsView (yid, yname, yage, yrating)
AS
SELECT sid, sname, age, rating
FROM Sailors
WHERE age < 18;

SELECT * FROM YoungSailorsView;

SELECT yname, yrating, yage
FROM YoungSailorsView
WHERE yrating >= 9;
```

**Another View Example (MySQL)**

```sql
CREATE VIEW ActiveSailors (sid, sname, rating)
AS
SELECT S.sid, S.sname, S.rating
FROM Sailors S WHERE EXISTS
(SELECT * FROM Reserves R WHERE R.sid = S.sid);

SELECT * FROM ActiveSailors;

UPDATE ActiveSailors
SET rating = 11
WHERE sid = 22;
```
So What About Views & Updates?

Ex:

CREATE VIEW SailsBoats AS
SELECT DISTINCT S.*, B.*
FROM Sailors S, Boats B, Reserves R
WHERE S.sid = R.sid and R.bid = B.bid;

Q: What if we now try...

UPDATE SailsBoats
SET rating = 12
WHERE sid = 22 AND bid = 101;

This view is not updatable since there is no update to the real (stored) tables that would have (just) the asked-for effect – see next slide...!

... Views & Updates (Cont’d.)?

UPDATE SailsBoats SET rating = 12
WHERE sid = 22 AND bid = 101;
... Views & Updates? (Cont’d.)

- A legal update $U$ to view $V$ must be translatable into an equivalent update $U'$ on the underlying table(s) $T$, i.e.:

```
Old Stored Table(s) T          Old View V
  apply update U'(T)           apply update U(V)
  compute view V(T)            compute view V(T')
New Stored Table(s) T'        New View V'
```

- If this isn’t possible, the system will reject the update
- Systems differ in how well they do this and err on the conservative side (i.e., declining more view updates)

SQL Access Control

- Based on the concept of access rights or privileges for objects (tables, views, stored procedures, ...) and mechanisms for giving users privileges (and revoking privileges).
- Creator of a database object automatically gets all privileges on it.
  - DBMS keeps track of who subsequently gains and loses privileges, and it ensures that only requests from users who have the necessary privileges (at the time the request is issued) are allowed to execute.
**GRANT Command**

```sql
GRANT privileges ON object TO users [WITH GRANT OPTION]
```

- The following privileges can be specified:
  - SELECT: Can read all columns (including those added later via ALTER TABLE command).
  - INSERT(col-name): Can insert tuples with non-null or non-default values in this column.
    - INSERT means same right with respect to all columns.
  - DELETE: Can delete tuples.
  - REFERENCES (col-name): Can define foreign keys (in other tables) that refer to this column.

- If a user has a privilege with the **GRANT OPTION**, can pass privilege on to other users (with or without passing on the **GRANT OPTION**).
- Only the owner can execute CREATE, ALTER, or DROP.

---

**GRANT and REVOKE of Privileges**

- GRANT  INSERT, SELECT ON Sailors TO Horatio
  - Horatio can query Sailors or insert tuples into it.
- GRANT DELETE ON Sailors TO Yuppy WITH GRANT OPTION
  - Yuppy can delete tuples and can authorize others to do so.
- GRANT UPDATE (rating) ON Sailors TO Dustin
  - Dustin can update (only) the rating field of Sailors tuples.
- GRANT SELECT ON ActiveSailors TO Guppy, Yuppy
  - This does NOT allow the ‘uppies to query Sailors directly!
- **REVOKE**: When a privilege is revoked from X, it is also revoked from all users who got it **solely** from X.
GRANT/REVOKE on Views

- Great combination to enforce restrictions on data visibility for various users/groups
- If view creator loses the SELECT privilege on an underlying table, the view is dropped!
- If view creator loses a privilege held with the grant option on an underlying table, (s)he loses it on the view as well – and so do users who were granted the privilege on the view!

Views & Security

- Views can be used to present just the necessary information (or a summary) while hiding some details of the underlying relation(s):
  - Given ActiveSailors, but not Sailors or Reserves, we can find sailors who have a reservation, but not the bid's of boats that have been reserved.
- Creator of a view has a privilege on the view if (s)he has the privilege on all underlying tables.
- Used together with GRANT/REVOKE commands, views are a very powerful access control tool.
- Stored procedures can be utilized similarly!
SQL Summary (I)

- SQL was a big factor in the early acceptance of the relational model; users found it more natural than earlier, procedural query languages.
- SQL is relationally complete; it has significantly more expressive power than the relational algebra.
- Queries that can be expressed in rel. alg. can often be expressed more naturally in SQL. (Ex: \texttt{max})
- Many alternative ways to write a query; optimizer will look for the most efficient evaluation plan.
  - In practice, expert users are aware of how queries are optimized and evaluated. (Optimizers are imperfect.)

SQL Summary (II)

- NULL for unknown field values brings many complications (as well as a SQL specification divergence for Oracle \texttt{w.r.t.} VARCHAR data).
- Allows specification of rich integrity constraints (real RDBMSs implement just some of SQL IC spec).
- Triggers can respond to changes in the database (and make up the difference when the set of available integrity features falls short).
- Stored procedures (and CALL) are also available.
- Views and authorization are both useful features, and can be especially powerful in combination. (!)
That’s it for SQL!

- ANY LINGERING QUESTIONS...?

Roadmap Check...

| Syllabus |
|-----------------|------------------|
| **Topic**       | **Reading**      |
| Databases and DB Systems | Ch. 1            |
| Entity-Relationship (E-R) Data Model | Ch. 2.1-2.5, 2.8 |
| Relational Data Model | Ch. 3.1-3.2      |
| E-R to Relational Translation | Ch. 3.5          |
| Relational Design Theory | Ch. 19.1-19.6, 20.8 |
| Midterm Exam 1   | Max. 30 (in class) |
| Relational Algebra | Ch. 4.1-4.2      |
| Relational Calculus | Ch. 4.3-4.4      |
| SQL Basics (SPJ and Nested Queries) | Ch. 3.4, 5.1-5.3 |
| SQL Analytics (Aggregation, Nulls, and Outer Joins) | Ch. 5.4-5.6 |
| Advanced SQL Goodies (Constraints, Triggers, Views, and Security) | Ch. 3.3, 3.6, 5.7-5.9, 21.1-21.3, 21.7 |
| Midterm Exam 2   | Wed, Apr 22 (in class) |
| Tree-Based Indexing | Ch. 8.1, 8.1-8.3, 10.1-10.2 |
| Hash-Based Indexing | Ch. 10.3-10.8, 11.1 |
| Physical DB Design | Ch. 8.5, 20.1-20.7 |
| Semistructured Data Management (a.k.a. NoSQL) | AsterixDB SQL++ Primer, Couchbase SQL++ Book |
| Basics of Transactions | Ch. 16 and Lecture Notes |
| Endterm Exam     | Fri, Jun 7 (in class) |
Disks and Files

- DBMSs store data on secondary storage.
- This has major implications for DBMS design!
  - **READ**: xfer data from disk (or flash) to memory (RAM).
  - **WRITE**: xfer data from RAM to disk (or flash).
  - Both are high-cost operations, relative to in-memory operations, so must be considered carefully!

Storage Hierarchy & Latency (Jim Gray in the mid 1990’s): How Far Away is my Data?

- $10^9$ Tape/Optical Robot: 2,000 years
- $10^6$ Disk: 2 years
- Memory: 1.5 hr
- On Board Cache: 10 min
- On Chip Cache: 1 min
- Registers: 1 min
- This Room
- My Head
- This Building
- La Jolla (UCSD)
- Andromeda
- Pluto
- 10
- 2
- 1
- 20
Why Not Store Data in Main Memory?

- **Main memory (RAM) costs too much!** Roughly:
  - RAM: $22/GB [Dell @ 05/2019]
  - SSD: $1/GB (22x cheaper than RAM)
  - Disk: $0.16/GB (138x cheaper than RAM, 5.5x vs. SSD)

- **Main memory is volatile.** We want our data to be saved between runs. (Obviously...!)

- Your typical (basic) storage hierarchy:
  - Main memory (RAM) for currently used data
  - (SSD often sits here now as well...)
  - Disk for the main database (secondary storage)
  - Tapes for archiving the data (tertiary storage)

Components of a Disk

- The platters spin (10,000 rpm)
- The arm assembly is moved in or out to position a head on a desired track
- Tracks under heads form a *cylinder* (imaginary!)
- Only one head reads/writes at any one time.
- **Block size** is a multiple of *sector size* (which is fixed)
Accessing a Disk Page

- Time to access (read/write) a disk block:
  - Seek time (moving arms to position disk head on track)
  - Rotational delay (waiting for block to rotate under head)
  - Transfer time (actually moving data to/from disk surface)

- Seek time and rotational delay dominate!
  - Seek time varies from about 1 to 20msec
  - Rotational delay varies from 0 to 10msec
  - Transfer rate is < 1 msec per 4KB page

- Key to lowering I/O cost: Reduce seek/rotation delays!  → Bottom line: Random vs. sequential I/O