CS101 Lecture 30:  
Databases and  
Data-Driven Applications  

Aaron Stevens  
23 November 2010  

Web Traffic - pct of Page Views  

Source: alexa.com, 11/22/2010
What You’ll Learn Today

– What is Facebook?
– How does Facebook generate your profile page?
– What is a database? Why is this useful?
– How does a database work?
– What is the Facebook platform?

Facebook: Usage Statistics

<table>
<thead>
<tr>
<th>People on Facebook</th>
<th>More than 500 million active users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50% of our active users log on to Facebook in any given day</td>
</tr>
<tr>
<td></td>
<td>Average user has 130 friends</td>
</tr>
<tr>
<td></td>
<td>People spend over 700 billion minutes per month on Facebook</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Activity on Facebook</th>
<th>There are over 900 million objects that people interact with (pages, groups, events and community pages)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average user is connected to 80 community pages, groups and events</td>
</tr>
<tr>
<td></td>
<td>Average user creates 90 pieces of content each month</td>
</tr>
<tr>
<td></td>
<td>More than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) shared each month.</td>
</tr>
</tbody>
</table>

November 2010
What is Facebook, anyway?

Core Applications
- Profile, friends, social network, inbox, photos, groups, events, …

Platform
- Enables developers to write their own Facebook applications.

Platform
- More than one million developers and entrepreneurs from more than 180 countries
- Every month, more than 70% of Facebook users engage with Platform applications
- More than 550,000 active applications currently on Facebook Platform
- More than one million websites have integrated with Facebook Platform
- More than 150 million people engage with Facebook on external websites every month
- Two-thirds of comScore’s U.S. Top 100 websites and half of comScore’s Global Top 100 websites have integrated with Facebook

How does Facebook generate your page?

Facebook is a data-driven application.

What kind of data?
- Profiles, Status, Friends, Groups, Friends, Events, Photos, etc.

All of this data is stored in a database.

What’s a database?
Storing Data: Main Memory

Main memory is used to store programs and other data that are currently in use.
- LOAD, STOR operations

Advantage of memory: short access times
- Read/write times in nanoseconds ($10^{-9}$ sec)

Disadvantages of memory:
- relatively expensive ($/byte$)
- “volatile”

Storing Data: Secondary Storage

Secondary storage is used to store data for later use (examples: disks, CD, DVD).
- Data is written from memory to disk.
- When needed, data is read back into memory.
Secondary Storage

Advantages of secondary storage:
– relatively inexpensive ($/byte)
– not “volatile”

Disadvantage of secondary storage: long access times
– Read times in milliseconds ($10^{-3}$ sec).
– in 10 ms, a modern CPU can perform millions of operations!

Thus, it’s important to minimize the number of times that the disk is accessed.

Example: Facebook Profile

What info is on a Facebook profile page?

- First Name
- Last Name
- Email
- Password
- Birthday
- About me
- Activities
- Favorite Books
- Favorite Movies
- Favorite Music
- Favorite Quotes
- Favorite TV shows

No matter which page you view, it has similar elements...

These data are stored in a database table.
What’s a Table?

Each table (sub-collection) is a collection of records, and each record contains fields. Example: a profiles table

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Email</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>Aaron</td>
<td><a href="mailto:ari@bu.edu">ari@bu.edu</a></td>
<td>Teaching, chasing Caleb down the street, home renovations, financial planning, investment</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>Anne</td>
<td><a href="mailto:anna@yahoo.com">anna@yahoo.com</a></td>
<td>Too much and vexed to list ...</td>
</tr>
<tr>
<td>3</td>
<td>L</td>
<td>Mike</td>
<td><a href="mailto:mikel@google.com">mikel@google.com</a></td>
<td>Hiking in the Metroarks, playing music, graphic design, home remodeling, playing with my dog</td>
</tr>
<tr>
<td>4</td>
<td>L</td>
<td>Leah</td>
<td><a href="mailto:leah@yahoo.com">leah@yahoo.com</a></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>S</td>
<td>Jennifer</td>
<td><a href="mailto:jen@ou.edu">jen@ou.edu</a></td>
<td>Michael Pollan: What a Guy. Please vote for Glen Gater for District Planning Board, Wes Welker</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>Josh</td>
<td><a href="mailto:josh@gmail.com">josh@gmail.com</a></td>
<td>Pretty liberal - let's be honest.</td>
</tr>
</tbody>
</table>

The primary key is a field which uniquely identifies one record within a table.
Consider this URL:

http://www.facebook.com/profile.php?id=919184

Primary Key

The Primary Key:
– uniquely identifies a record within a table
– is an ideal search key
– is a way to create relationships between different tables

Consider this URL again...

http://www.facebook.com/profile.php?id=919184
Example: Status Updates

Not all data fits neatly into the profile table. Consider status updates...

A separate Status table tracks all status updates for all users.

What fields does it need?
- DateTime, ID, Message

Example: Status Table

Each status message is related to exactly one profile by the foreign key (the field called 'id').

<table>
<thead>
<tr>
<th>DateTime</th>
<th>ID</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-02-27 20:39:00</td>
<td>3</td>
<td>Watching 'Move Than A Game'. Great movie!</td>
</tr>
<tr>
<td>2010-03-06 15:22:00</td>
<td>2</td>
<td>It's too pretty out for cooking and cleaning. Gonna go enjoy that lovely lake eventually.</td>
</tr>
<tr>
<td>2010-03-10 19:23:00</td>
<td>3</td>
<td>Yoni loves to listen to Jewell's version of Twinkle Twinkle Little Star as we get ready for bed. That's funny 'cause it's a star, it's a, it's a star...</td>
</tr>
<tr>
<td>2010-03-11 08:30:00</td>
<td>2</td>
<td>Isn't up for it. Whatever it is.</td>
</tr>
<tr>
<td>2010-03-12 15:57:00</td>
<td>4</td>
<td>is a crabby lady today. Perhaps a nap will help.</td>
</tr>
<tr>
<td>2010-03-13 07:27:00</td>
<td>1</td>
<td>According to fox news, I'm a communist rice eater, and I should be stripped of my citizenship and deported to N Korea.</td>
</tr>
<tr>
<td>2010-03-13 12:44:00</td>
<td>4</td>
<td>is trying to take a nap but Jonah is loudly singing. Frogs here, frogs there, frogs are bumping everywhere. Yay!</td>
</tr>
<tr>
<td>2010-03-13 13:44:00</td>
<td>2</td>
<td>needs to stop brainless web searching and get something done today.</td>
</tr>
<tr>
<td>2010-03-13 16:09:00</td>
<td>3</td>
<td>Feed dog-check, Feed fish-check. Someone to watch Yoni-check. Double date with wife and friends before B-Day.</td>
</tr>
<tr>
<td>2010-03-13 23:17:00</td>
<td>4</td>
<td>is now down to only one pair of shoes that she can get onto her swollen feet. Just a little bit longer...</td>
</tr>
</tbody>
</table>

These ids are called a foreign key, because they are primary keys in another table.
Example: Mini FB Database

Let’s consider just 2 tables:

<table>
<thead>
<tr>
<th>Table 'profiles':</th>
<th>Table 'status':</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field</td>
<td>Type</td>
</tr>
<tr>
<td>ID</td>
<td>int</td>
</tr>
<tr>
<td>LastName</td>
<td>text</td>
</tr>
<tr>
<td>FirstName</td>
<td>text</td>
</tr>
<tr>
<td>Email</td>
<td>text</td>
</tr>
</tbody>
</table>

Notice that these tables share some data:
- ID is a primary key in profiles, and a foreign key in status; this enables cross-table relationships.

Example: Friends

The Facebook “friend” relationship is created by an entry in a friend table.

Each record has two user ids:

<table>
<thead>
<tr>
<th>ID1</th>
<th>ID2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

These ids are called a foreign keys, because they are primary keys in another table.
Databases

A database is:
– a collection of data stored in a way to enable quick access
– organized into related sub-collections called tables.

Example: a mini Facebook database:
• A profiles table has records of each user
• A status table has records of status messages
• A friends table has records of friend relationships

Each table (sub-collection) is organized by records, and each record contains fields.

Database Management Systems

A database is a collection of data (not software). A database management system (DBMS) is the software which manages a database.

Functions of a DBMS:
Efficient storage
Providing a logical view of data (tables, records)
Query processing
Transaction management
Database Management Systems

Efficiency - Indexing enables locating a record based on some unique attribute, called a key.
   – Example: looking up profiles by their id.

Logical representation - storage by record.
   – Example: update the record for ID='5'.

Query Processing

A query language is used to access and modify the data.

SQL (Structured Query Language) is the standard for relational databases.
   Many different database vendors support SQL:
      – Oracle, Sybase, IBM DB2, MS SQL Server, MS Access
      – MySQL, SQLite (free/open-source)
Transaction Integrity

A transaction is an atomic sequence of operations.
- Example: booking a flight
  - Select flight, reserve seat, make payment

DBMS guarantees transaction integrity: completion or failure.

Database Applications

End users rarely interact with a database directly.

A database-enabled application allows the users to interact with the database without needing to know the query language.
SQL Example

The **SELECT** query can specify conditions using the **WHERE** clause, which creates a more refined result set (e.g. only matching records are returned).

Example:

```sql
SELECT *
FROM profiles
WHERE ID=2
```

<table>
<thead>
<tr>
<th>ID</th>
<th>LastName</th>
<th>FirstName</th>
<th>Email</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>S</td>
<td>Anna</td>
<td><a href="mailto:annas@yahoo.com">annas@yahoo.com</a></td>
<td>Too much and varied to list . . .</td>
</tr>
</tbody>
</table>
Facebook platform

Development tools to allow programmers to develop applications for the Facebook website.
- Web services API
- Database query language (FQL)
- Text markup languages (FML)

Why create a Facebook Platform?
Isn’t this giving away the store?

Facebook: Ethical Questions

- Public profile pages
- Privacy of messages
- Use in investigations
- No mechanism to “delete” your profile
- Workplace productivity
How to Make Money…

– What is the business of Facebook?
– Who are Facebook’s customers?
– What are Facebook’s assets?

Facebook Haiku

AVOIDING FACEBOOK
BECAME TOO MUCH OF A CHORE
SO I RELENTED

– SCOTT M. SOKOL
What You Learned Today

– Database, DBMS
– Data-driven application
– Tables, Records, Fields
– Primary Key, Foreign Key
– Structured Query Language
– Facebook platform

Student To Dos

– HW11 is due TUE 11/23
– THU 11/25 is Thanksgiving Holiday - BU CLOSED

– Southpark episode about facebook/social network: