# Dude, Where's my Database?

Eric Florenzano (@ericflo)

# Motivation

- Lots of databases
- Lots of hand waving
- NoSQL?
- One size doesn't fit all

# Categories

- Relational
- Key/Value
- Data Structure
- Graph
- Document-Oriented
- Highly Distributed

# Relational

- Highly structured
- Strong type system
- Powerful query language

# Relational

- PostgreSQL
- MySQL
- MSSQL
- Oracle
- Drizzle
- etc.

# Relational

+ Easy to use
+ Easy to conceptualize
+ Well-understood
+ Fits many problem domains
- Hard to scale

# Key/Value

- Unstructured
- No type system
- Extremely simplistic query API

# Key/Value

- gdbm (anydbm -> dbm in Python 3)
- Tokyo Cabinet/Kyoto Cabinet
- Berkeley DB
- MemcacheDB

# Key/Value

+ Simple
+ Fast
- No interesting queries

# Key/Value

- Tracking HTTP Sessions
- User Preferences
- URL Shorteners

# Data Structure

- Modification of Key/Value
- Structured Values
- Atomic Operations
- Redis

# Data Structure

+ Fast
+ Maps to certain problems very well
- Lack of alternative implementations
- Everything has to fit in memory*

# Data Structure

- Pageview Counter
- Task Queue
- Trend Analysis
- “... anything involving statistics or high volumes of small writes.” - Simon Willison

# Graph

- Store data as nodes and edges in a graph
- Fits logically to many problem spaces
- Programmatic queries

# Graph

- Neo4j
- VertexDB

# Graph

+ Maps logically to many problem spaces
+ Can do interesting queries because it's a graph
- Scale ceiling
- Lack of alternative implementations

# Graph

- Social Graph
- Threaded Comments
- Group Membership

# Document-Oriented

- Unstructured
- Formatted (JSON, Python Object)
- Programmable Query API

# Document-Oriented

- CouchDB
- MongoDB
- ZODB

# Document-Oriented

+ Documents/Objects map to real world
+ Queryable in interesting ways
- Scale ceiling
- Implementation-specific weaknesses

# Document-Oriented

- Activity Streams/Lifestreams
- User Data
- CMS

# Highly Distributed

- Optimized for multi-node
- Add and remove nodes on the fly
- Hard to do ad-hoc queries
- Sacrifice consistency

# Highly Distributed

- Cassandra
- Riak
- HBase
- Hypertable

# Highly Distributed

+ Very very high scale ceiling
+ Highly available
+ Eventual consistency
- Eventual consistency
- Can’t do efficient ad-hoc queries

# Highly Distributed

- Digg
- Google
- Yahoo!
- Mochi Media - We're hiring :)
  - bit.ly/mochi-jobs

# More Information

- NoSQL - Mark Ramm - 1:35pm
- MongoDB - Rick Copeland - 4:15pm
- Neo4j - Tobias Ivarsson - 4:55pm
- Cassandra open space - 5:30pm
- Make more open spaces!

# Questions?