

# Course Review

CPS352: Database Systems

Simon Miner  
Gordon College  
Last Revised: 5/6/15

# Agenda

- Check-in
- NoSQL/Data Science Technology Presentations
- Course Review
- Exam II

# Check-in

# NoSQL/Data Science Technology Presentations

# Course Review

# Modeling in the Logical Layer

- Relational model
  - Entities connected by relationships
  - Keys and nulls
  - Schema Diagrams
- Relational algebra
  - Selection
  - Projection
  - Joins – Cartesian/natural/theta
  - Set operators – union/difference/intersection
  - Rename
  - Outer join
  - Semi-join and anti-join
  - Aggregate functions
- Entity relationship model
  - Entities and entity sets
  - Relationships and relationship sets
  - Attributes – atomic/composite/derived/multi-valued
- Mapping cardinalities
  - One-to-one
  - One-to-many
  - Many-to-many
- Total participation constraint
- Weak entities
- Generalized and specialized entities

# SQL

- DML
  - Select
    - Joins
    - Group by / having
    - Order by
    - Subqueries
    - Recursive queries
  - Insert
  - Update
  - Delete
  - Commit / rollback
- DDL
  - Create/alter/drop table
  - Create view
  - Create index
- Integrity constraints
  - Primary key/unique
  - Foreign key (referential)
  - Domain constraints
  - Check clause
  - Triggers
- Security
  - User accounts
  - Grant/revoke statements

# Database Design Theory and Normalization

- Functional dependencies (FD)
  - Closure
  - Canonical cover
  - Super/candidate/primary key
- Multi-valued dependencies (MVD)
- Decomposition
  - Lossless join
  - Dependency preserving
- Database design goals
  - Avoid redundancies
  - Ensure lossless join
  - Ensure dependency preserving decompositions
- Normal forms
  - 1NF – atomic
  - 2NF – no partial key dependencies
  - 3NF – no transitive dependencies
  - BCNF – key, whole key, and nothing but the key
  - 4NF – BCNF + non-redundant MVDs



# Database Application Development

- Evolution of database clients (thick -> thin -> web)
- Database access from applications –
  - Dynamic (JDBC-style)
  - Static/embedded (SQLJ)
  - Object relational mapping
- Application architecture
  - Two- vs. three-tier models
  - MVC – Model (+business) / View / Controller layers
- Database Design Tips
  - Choose good names – succinct yet clear
  - Columns every table should have – id, created\_at, updated\_at, status
  - Keep business logic out of presentation and data access layers

# Database Physical Layer

- Minimize disk accesses
  - Storage system setup
    - Disk vs. memory buffer
    - RAID levels
- Record organization
  - Fixed vs. variable length
  - Sequential vs. multi-table clustering
- Indexing
  - Ordered vs. hashed
  - Clustered vs. non-clustered
  - Dense vs. sparse
  - B+ tree indexes (and searches)
- Query optimization
  - Selection – full table, index based on type (exact vs. range)
  - Join strategies
    - Nested loop
    - Nested block
    - Buffering entire relation
    - Merge join
  - Ordering joins + equivalence rules
    - Push selections inward
    - Push projections outward
  - Estimating join size with statistics

# Concurrency

- ACID
- Transactions
  - Transaction states
  - Schedules
  - Serializability
  - Precedence graphs
  - Recovery – cascading rollback
- Crash recovery
  - Transaction log
  - Approaches
    - Incremental log with deferred update
    - Incremental log with immediate update
    - Shadow paging
- Locking
  - Granularity of locks
  - Shared vs. exclusive locks
  - Deadlock
  - Two-phase locking protocol
    - Growth and shrinking phases
- Other concurrency approaches
  - Timestamps
  - Validation – optimistic concurrency
  - Multiversion schemes
- Inserts, deletes and phantom rows
- Relaxing consistency

# Database Architectures

- Parallelism
  - Speed-up vs. scale-up – batch vs. transaction scale-up
- Distributed systems
  - Fragmentation (horizontal vs. vertical)
  - Replication
  - Data transparency
  - Two-phase commit protocol
  - Concurrency issues (locking, timestamps)
- NoSQL...

# NoSQL

- Why NoSQL?
- Common characteristics
- Aggregate-oriented databases
- Schema-less databases
- Scaling vs. consistency
  - Sharding and replication
  - Update and read consistency
- Data models
  - Key-value databases
  - Document databases
  - Column-family databases
  - Graph databases
- Map-reduce pattern
- Schema migrations
- Polyglot persistence
- When (not) to use NoSQL

# Data Analysis

- OLTP vs. OLAP
- Data warehouses
- OLAP concepts
  - Dimension vs. measurement attributes
  - Cube
  - Rollup
  - Rank and dense rank
- Data Science
  - Data wrangling/munging
  - Analysis
  - Communication (visualizations)

# Exam II