

Database Management Systems

Final Review

Learning Outcomes

- Be able to write syntactically and logically correct and easy to understand queries using SQL, Relational Algebra and Datalog;
- Be able to develop database applications to solve problems;
- Be able to design good database schema or convert existing database schema to BC Normal Form, and use SQL to realize the schema and related constraints;
- Understand the database security concepts and be able to use SQL to manage the database security;
- Have some understanding about the internal working of a database engine, such as transaction management, recovery and index building.

Query Languages

- Relational Algebra
 - Operands: relations
 - Operators: selection, projection, join, set union, set intersection and set difference
- Datalog
 - Declarative language
 - Unsafe queries
- SQL
 - Data Definition Language
 - Data Control Language
 - Data Manipulation Language

SQL

- SQL update
 - insert (one literal valued tuple or multiple derived tuples), delete, update
- SQL query
 - 6 clauses: select, from, where, group by, having, order by
- SQL sub-queries
 - column IN/NOT IN (set-generating-query)
 - EXISTS/NOT EXISTS (table-generating-query)
 - column Relational-Operator (=, <>, >, <, >=, <=) SOME/ALL (set-generating-query)
 - table-generating-query in the FROM list
- Joins in SQL: cross product with where condition, join/on condition, natural join, outer (left/right/full) join
- Aggregation: functions, group by requirement
- Null values: how they are tested, how they are treated in aggregation functions

Database Applications

- Principle: As few customized code as possible
- client-server architecture
- connect to the database
- execute (updating or retrieving) query
- access the returned data (result set)
- properly close connection and clean up the environment
- prepared (parameterized) statements, setting parameters, and the advantages in using prepared statement
- unprepared statements, and its potential security problems if not handled properly

Database Design

- Domain knowledge analysis
- ER Diagram
 - strong (regular)/weak entity sets
 - relationship sets (cardinalities: M-1, M-M)
 - attributes (identifiers/discriminators)
 - aggregation (treat a relationship set as an entity set)
 - inheritance (specialization, generalization)
- Translate ER Diagram to Relational Schema
- Where do the foreign keys come from?

Database Design Theory

- Functional Dependency
- Reasoning of FD
- Keys: super keys, candidate keys, and primary keys
- Decomposition
 - lossless join decomposition
 - dependency preserving decomposition
- BC Normal Forms
 - Definition
 - process of decomposing a relation to BC normal form
- Third Normal Forms
 - Definition

Security, etc

- Security
 - privileges
 - grant/revoke SQL statement
- View
 - concept
 - create/drop view
- Constraints
 - Primary key
 - Foreign key
 - General check constraints

Transaction Management and Recovery

- ACID properties: Atomicity, Consistency, Isolation and Durability
- Anomalies: read dirty data (WR), unrepeatable reads (RW), lost write (WW), and phantom problem
- Serial schedules: one by one
- Conflict serializable schedules: conflict equivalent to a serial schedule
 - Definition of conflict operations
 - Serialization (precedence) graph
- Recoverable and cascade-less schedules
- Two Phase Locking and Strict Two Phase Locking protocols
 - Deadlock and how to prevent/recover
- Timestamp based protocols
- Database Recovery
 - Write Ahead Logging Protocol
 - ARIES algorithm

Indexing

- Data on disk, database storage model
- B+ tree index
 - Primary (sparse, clustered) index
 - Secondary (dense, non-clustered) index
- How to use B+ tree index in processing queries
- Hash table index