同济大学 TONGJI UNIVERSITY

ADMIS Lab
先进数据与机器智能系统实验室

Database System Principles
《数据库系统原理》

关佶红/Jihong Guan
教授/Professor
Email: jhguan@tongji.edu.cn

张毅超/Yichao Zhang
副教授/Associte Professor
Email: yichaozhang@tongji.edu.cn

李文根/Wengen Li
副教授/Associte Professor
Email: lwengen@tongji.edu.cn

同济大学 TONGJI UNIVERSITY
电子与信息工程学院 College of Electronics and Information Engineering
计算机科学与技术系 Department of Computer Science and Technology

课程概要

同济大学 TONGJI UNIVERSITY

- **Part 0: Overview**
  - Ch1: Introduction
- **Part 1 Relational Databases**
  - Ch2: Relational model (data model, relational algebra)
  - Ch3&4: SQL(Structured Query Language)
  - Ch5: Advanced SQL
- **Part 2 Database Design**
  - Ch6: Database design based on E-R model
  - Ch7: Relational database design
- **Part 3 Application Design & Development**
  - Ch8: Complex data types
  - Ch9: Application development
- **Part 4 Big data analytics**
  - Ch10: Big data
  - Ch11: Data analytics
- **Part 5 Data Storage & Indexing**
  - Ch12: Physical storage system
  - Ch13: Data storage structure
  - Ch14: Indexing
- **Part 6 Query Processing & Optimization**
  - Ch15: Query processing
  - Ch16: Query optimization
- **Part 7 Transaction Management**
  - Ch17: Transactions
  - Ch18: Concurrency control
  - Ch19: Recovery system
- **Part 8 Parallel & Distributed Database**
  - Ch20: Database system architecture
  - Ch21-23: Parallel & distributed storage, query processing & transaction processing
- **Part 9**
  - DB Platform: **OceanBase**, MongoDB, Neo4J

Chapter 1 Introduction

- **Drawbacks of file-processing system**
  - Data redundancy and inconsistency
  - Difficulty in accessing data
  - Data isolation
  - Integrity problems
  - Atomicity of updates
  - Concurrent access by multiple users
  - Security problems
- Database, Database Management System (DBMS), Database system
- **Abstraction for database**
  - Physical level, logical level, view level
- Schema vs. instance
- DB conceptual/logical design
- **DB users**
  - Different user types
  - The duties of DBA
- Major DMBS systems

Chapter 2 Relational Model

- **Concepts**
  - Relational schema, relation, relation instance, tuple, attribute
  - Superkey, candidate key, primary key, foreign key, referencing relation, referenced relation
  - Query language, procedural, non-procedural, relational algebra
- **Relation algebra**
  - Six basic operators of relational algebra: select, project, rename, union, set difference, cartesian product
  - Set intersection, natural join, assignment
  - Generalized projection, outer join, aggregate functions
  - Null values
  - Deletion, insertion, updating
  - Views

Chapters 3, 4 & 5 SQL

- **Components of SQL**
  - DDL & DML
- **Domain types in SQL**
- **Create table construct**
  - Integrity constraints
  - Insertion and deletion
  - Drop and alter
- select, rename, ordering of tuples, duplicates, set operations, aggregation
- Null values
- Nested subqueries
- Set comparison
- Views
- **Modification of the Database**
  - Deletion
  - Insertion
  - Updates

Chapters 3, 4 & 5 SQL (续)

同济大学 TONGJI UNIVERSITY

- **Integrity constraints**
  - Domain constrains, unique constrains, referential integrity, check, assertion
- **Security and authorization**
  - Levels of security (DB, OS and Network)
  - Forms of authorization to modify DB
  - Granting privileges, role
  - Audit trails, encryption, authentication
- Embedded SQL
  - Update through cursors
- Dynamic SQL
- Functions and procedural constructs
- Recursive queries
- Advanced SQL features

Chapter 6 Database Design and E-R Model

- **Database Design**
  - Conceptual design
  - Logical design
  - Physical design
- **Relationship set**
  - Degree of mapping cardinality
- **ER design issues**
  - use of entity sets vs. attributes
  - use of entity sets vs. relationship sets
  - binary vs. n-ary relationship sets
  - placement of relationship attributes
  - use of specialization/generalization
- Symbols used in E-R diagram
- Reduction of an E-R diagram to relational tables

Chapter 7 Relational Database Design

- **Functional Dependency/函数依赖**
  - Closure of functional dependency
    - Armstrong's axioms: reflexivity(自反), augmentation(增广), transitivity(传递)
    - Additional rules: union(合并), decomposition(分解), pseudotransitivity(伪传递)
  - Closure of attribute set
  - Canonical cover
    - Extraneous attribute and its testing
  - Find all candidate keys
- **Decomposition**
  - Lossless-join decomposition & dependency preservation
- **Normalization & Normal Forms (范式)**
  - Normalization: 1NF, 2NF, 3NF, BCNF
  - Testing for BCNF and 3NF, and the corresponding decomposition algorithms

Chapters 12-13 Storage and File Structure

- **Classification of physical storage media**
  - Speed, cost, reliability, volatile, non-volatile
  - Cache, main memory, flash, magnetic disk, optical storage, tape storage
- **Performance measure and optimization of disk access**
  - Access time, data transfer rate, mean time to failure
  - Block, disk-arm-scheduling
- **Storage access and buffer manage**
- File organization
  - Fixed-length, variable-length records, organization of records in files
- Data dictionary storage
- Raid levels 1-6

Chapter 14 Indexing

- **Basic concepts**
  - Search key and index file, ordered index and hashed index
  - Equal query and range query
  - Primary index or clustering index
  - Secondary index or non-clustering index
  - Dense, sparse, multilevel indices
- **Ordered index**
  - B^+-tree index, B-tree index
- **Hash index**
  - Static hashing
  - Dynamic hashing: extendable hashing
- **Ordered indexing vs. hashing**
- **Index definition in SQL**
- Multiple-key access: grid files and bitmap index

Chapter 15 Query Processing

- **Basic steps in query processing**
  - Parsing and translation, optimization, evaluation
- **Measures of query cost**
  - Seek time, number of data transfers
- **Selection operation**
  - File scan vs. index scan
  - A1-A10
- **Sorting**
  - External sort-merge

Chapter 15 Query Processing

- **Join operation**
  - Nested-loop join
  - Block nested-loop join
  - Indexed nested-loop join
  - Merge-join
  - Hash-join
- **Other operations**
  - Duplicate elimination, aggregation, set operations, outer join,
- **Evaluation of expressions**
  - Materialization(物化)
  - Pipelining(流水线)

- **Evaluating a given query**
  - Equivalent expressions: equivalence results
  - Different algorithms: measure of cost
- **Measures of cost**
  - Select size
  - Join size
  - Distinct values
- Evaluation plan
- Practical query optimizers combine the following two approaches
  - Search all the plans and choose the best plan in a cost-based fashion: dynamic programming
  - Use heuristics to choose a plan

Chapter 17 Transactions

- **Transaction**
  - What's transaction and the two main issues to deal with?
  - Properties: ACID
  - Transaction state
  - Concurrent execution
- **Serializability**
  - Conflict serializability
  - View serializability
- Recoverability, cascading rollback, cascadeless schedule
- **Testing for Serializability**
  - Precedence graph

Chapter 18 Concurrency Control

- **Lock-based protocols**
  - Lock-compatibility matrix
  - Deadlock and starvation
  - Two-phase locking protocol
    - Strict two-phase locking
    - Rigorous two-phase locking
  - Lock manager and lock table
  - Graph based protocol: impose a partial ordering
  - Timestamp-based protocols
- **Multiple granularity**
  - Intention lock modes
  - Compatibility matrix with intention lock modes
- **Deadlock handling**
  - Deadlock prevention protocols
    - graph-based protocol
    - schemes use transaction timestamps: wait-die scheme (非抢占) and wound-wait (抢占)
    - timeout-based schemes
  - Deadlock detection
    - wait-for graph

Chapter 19 Recovery System

- **Two approaches for recovery**
  - log-based recovery
    - Deferred database modification
    - Immediate database modification
    - Checkpoint
- **Recovery with concurrent transactions**