

Database Development and Design (CPT201)

Lecture 14: Revision

Dr. Wei Wang
Department of Computing

About the Exam

- Answer **ALL FOUR** questions
 - Each carries 25 marks
 - Open book exam, but NO electronic devices
 - 100% of final assessment (resit will also be 100%)
 - Time allowed: 2 hours (for online exam there will be 30 minutes crash hour *subject to registry approval*)
- Difficulty level
 - Comparable to the ones in tutorials
 - Comparable to the past year papers
- Calculators allowed by the university

About the Online Exam

- DO NOT EMAIL YOUR ANSWERS! THEY WILL NOT BE READ!
- IF YOU HAVE ANY PROBLEMS IN SUBMISSION, CONTACT THE REGISTRY IMMEDIATELY AT Registry@xjtlu.edu.cn
- SUBMIT VIA LINK ON LEARNING MALL!
- SUBMISSION LINK WILL BE DEACTIVATED AFTER THE CRASH HOUR! THIS IS CONTROLLED BY THE SYSTEM!

Overview of Exam Structures

- Question 1: Indexing, B+tree index, hash index
- Question 2: Query evaluation and optimisation
- Question 3: Transaction management, concurrency, and failure recovery
- Question 4: Market basket analysis, DDBMS, OODB, XML, RDF, noSQL

NOT necessarily in the exact order

Q1: Indexing

- Concepts about ordered index, hashing index, clustering index, non-clustering index, primary index, secondary index, dense index, sparse index
- B+ tree indexing
 - Properties (e.g., validness, height)
 - Insertion
 - Deletion
- Hash indexing
 - Insertion in extendable hash index (deletion not required)
 - Use of overflow buckets

Q2: Query Evaluation and Optimisation

- Selection evaluation cost
 - Linear search and binary search
 - Selection using index (primary or secondary index, dense or sparse index, etc)
 - Comparative selection
- External merge sort algorithm (number of sorted run files, merge passes, block transfer and seeks)
- Join evaluation cost in terms of number of block transfers and seeks
 - Nested loop
 - Block nested loop
 - Indexed nested loop
 - Merge join
 - Hash join

Q2: Query Evaluation and Optimisation cont'd

- Query evaluation plan (usually in a form of diagrams) and relational algebra expression
- Relational algebra expression transformation using equivalence rules (only the simple ones)
- Heuristics for query optimisation, e.g., push selection and push projections
- Size estimation
 - Size of selection
 - Size of join
- Cost estimation, usually combined with query evaluation
 - If pipelining is used, no need for block transfers and seeks
 - Cost estimation for the whole evaluation plan

Q3: Transaction/concurrency/recovery

- Concept of conflict serialisable schedule, precedence graph
- Recoverable schedule and cascadeless schedule
- View serialisability
- Deadlock
 - Deadlock detection, e.g., using wait-for graph
 - Deadlock prevention based on timestamp: wait-die and wound-wait
 - Deadlock recovery strategies, starvation
- Two-phase locking protocol/strict and rigorous 2PL protocols
- Log-based recovery
 - Algorithm (redo and undo)
 - Logs for recovery
 - Checkpoint

Q4: Market basket analysis, DDBMS, OODB, XML, RDF, noSQL

- Market basket analysis and association rules
 - Support and confidence
- DDBMS
 - Storing data, distributed join (semi-join and bloom join)
 - Centralised deadlock detection method
 - Two-phase commit protocol
 - Failure recovery: coordinator or participating sites failures
- OODB: data representation and modelling
- XML/RDF: data representation and modelling
- noSQL and big data
 - Eventual consistency
 - Different types of noSQL/big data stores

GOOD LUCK!