



CSE594: REVIEW

1



CSE594 in one slide

- ❖ **Logical:** E/R diagram → normalized relations
- ❖ **Physical:** files, buffering, and indexes
- ❖ **Logical:** Relational algebra and SQL
- ❖ **Physical:** join methods, optimization
- ❖ **Logical and physical:** DBA designs, tunes
- ❖ **CC/R** choices affect correctness, performance
- ❖ **Object Databases** extend, improve relations
 - ... OODBMS, ORDBMS are converging?
 - ... OQL, SQL3 are converging?

2



The E/R model

- ❖ Entities vs. relationships
- ❖ Attributes for entities and relationships
- ❖ Keys and weak entities
- ❖ Cardinality constraints
- ❖ Participation (key) constraints
- ❖ Translation to relations:
 - Entity becomes relation
 - Relationship becomes relation iff it is many-many
 - Other relationships: key of "parent" goes with "child"

3



FDs and Normal Forms

- ❖ Determine Candidate Keys (CKs)
- ❖ A field is prime iff it is in some CK
- ❖ 1NF: All relations are in 1NF
- ❖ 2NF: No FDs $X \rightarrow A$ where $X \subset$ some CK and A is non-prime
- ❖ 3NF: For each FD $X \rightarrow A$:
 - X is a superkey OR A is prime OR $A \in X$ (trivial)
- ❖ BCNF: For each FD $X \rightarrow A$:
 - X is a superkey OR $A \in X$ (trivial)
- ❖ Normalize by decomposing R (ABC) to R1(AB) and R2 (BC), where $B \rightarrow C$ violates a NF

4



Physical DB Design

- ❖ File organizations: heapfile, sorted file, hashed file
- ❖ Disks: speed and physical characteristics
- ❖ Buffer pool can vary in size; frame holds page
- ❖ Replacement policy choice is important
- ❖ Indexes can speed up data access:
 - Clustered vs. unclustered
 - Primary vs. secondary
 - Dense vs. sparse
 - B+ Trees are the most common: $\log(N)$ searches

5



Queries: Logical Aspects

- ❖ Relational Algebra
 - $\sigma, \pi, \times, \cup, -$
 - Joins kind of important too!!!
 - ♦ General join
 - ♦ Equijoin
 - ♦ Natural join
 - Transformations for optimization
- ❖ SQL
 - Declarative
 - Relational calculus + algebra + grouping, etc.
 - Computationally incomplete
 - Query execution: relational algebra operations

6

Queries: Physical Aspects

- ❖ Implementation of relational algebra
 - File scan vs. index lookup vs. binary search
 - Exact-match vs. range queries
 - Impact of clustered vs. non-clustered indices
 - Join methods:
 - ♦ Tuple-oriented nested loops (dumb idea)
 - ♦ Page-oriented nested loops (decent idea)
 - ♦ Block-oriented nested loops (better idea, if buffers avail.)
- ❖ Query optimization: cost estimates hard!!!
 - Large space of physical, logical alternatives
 - Prune space by considering only left-deep plans
 - Enables pipelined execution

7

Tuning

- ❖ Ongoing task of DBA:
 - Monitor performance
 - Adjust physical schema as necessary (indices, clusters, etc.)
 - ♦ Physical data independence is preserved!!
 - Adjust logical schema as necessary
 - ♦ Hide changes using views
 - ♦ Maintain integrity constraints using triggers, etc.
 - ♦ Logical data independence is preserved!!

8

Concurrency Control and Recovery

- ❖ ACID transactions enforce CC&R
- ❖ Strict 2PL guarantees serializability (C, I)
 - Deadlocks possible (detect and abort a XACT)
 - Phantom tuples possible (fix with index locking)
- ❖ ARIES guarantees XACTS are A, D
 - Write-ahead logging
 - Log UNDO actions to allow complete replaying of history
 - Recovery phases:
 - ♦ Analyze (rebuild main memory)
 - ♦ REDO (rebuild disks)
 - ♦ UNDO (abort incomplete XACTS)

9

Object Databases

- ❖ ORDBMS (e.g. Oracle 8i) vs. OODBMS (e.g. O2)
- ❖ Structural aspects
 - Collection-valued fields
 - OIDs
 - Freely composable type constructors
- ❖ Behavioral aspects (methods)
- ❖ Adds new dimensions to RDBMS problems
- ❖ SQL3, OQL standards are emerging

10