

CSE 344: Lecture 26

Final Review

Guest Lecturer: Paris Koutris

The Final

- Wednesday, March 14th, 8:30-10:30
- In class

The Final

Questions:

1. SQL, Relational Calculus, Relational Algebra(lectures 2-10)
2. XML/XPath/XQuery (lectures 11-12)
3. E/R diagrams, constraints, conceptual design (lectures 14-16)
4. Transactions (lectures 18-20)
5. Parallel data processing (lectures 21-24)
6. DBMSs as a service (lecture 25)

1. SQL, Relational Calculus, Relational Algebra

SQL

- SELECT-FROM-WHERE
- DISTINCT, ORDER BY, renaming of attributes
- INSERT, DELETE, UPDATE
- GROUP-BY and HAVING: *different* from WHERE (why ?)
- NULLs, outer joins
- Nested queries (subqueries)

Know the elements of the syntax

Know the semantics (nested loops !)

1. SQL, Relational Calculus, Relational Algebra

Relational Calculus, Relational Algebra

- Understand the existential/universal quantifiers
- Review RA and the extended RA
- Translation from SQL to RA

2. XML/XPath/XQuery

- XML
 - Basic definitions: tags/elements/attributes/text, well-formed/valid XML document
 - DTDs
- XPath – really easy...
- XQuery – a kind of SQL

3. E/R Diagrams, Constraints, Conceptual Design

E/R Diagrams

- Entities, attributes
- Relationships:
 - Many-many, many-one, one-one
 - Multi-way relationships
- Inheritance, weak entity sets, union types
- Constraints in E/R diagrams
- Translation to relations

3. E/R Diagrams, Constraints, Conceptual Design

Constraints in SQL

- Keys and Foreign Keys
- Attribute level constraints
 - Predicates on values
 - NOT NULL
- General constraints

3. E/R Diagrams, Constraints, Conceptual Design

Conceptual Design

- Data anomalies
- Functional dependencies
 - Definition
 - Make sure you can check if a table satisfies a set of FDs
- Attribute closure
- Keys and Super keys
- Definition of BCNF
- Decomposition to BCNF

4. Transactions

Transactions concepts

- Review ACID properties
- Definition of *serializability*
- The four isolation levels in SQL
- Concurrency control using locks
 - SQLite and SQLServer examples
- Phantoms, dirty reads, and other problems
- Deadlocks
- Transactions in SQL

5. Parallel Data Processing

- Parallel databases:
 - Speedup/scaleup
 - Shared memory, shared disk, shared nothing
 - How to implement simple algorithms: group-by, join
- MapReduce
 - Functions: map, (combine,) reduce
 - Terminology: map job / reduce job; map task / reduce task; server (instance)
 - Basic implementation of MR
 - Dealing with server failures and stragglers
- Pig system and Pig Latin language

6. DBMS-as-a-service

- Challenges and benefits
- Types of data management systems offered as services
- NoSQL systems
 - Motivation
 - Similarity and differences