CSE 344: Lecture 28

Final Review
The Final

• Wednesday, December 14th, 8:30-10:30

• In class

• Open notes and open books

• Special office hours: Monday December 12, 10-12pm, CSE 584
The Final

Entire class content is on the final!

But focus of questions on the final will be as follows:

1. E/R diagrams and constraints (lectures 14-15)
2. Conceptual database design (lecture 16)
3. SQL including views (lectures 1-7 and lecture 17)
4. Transactions (lectures 18-20)
5. Parallel data processing (lectures 21-24)
6. DBMSs as a service (lecture 25)
7. Data integration and data cleaning (lectures 26 and 27)
8. Datalog and relational algebra (lectures 8-10)
9. XML/XPath/XQuery (lectures 12-13)
1. E/R Diagrams and Constraints

- 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
- Constraints in SQL including triggers
2. Conceptual Design

• Normal forms and functional dependencies:
  – Anomalies
  – Functional dependencies
  – Attribute closures (X+)
  – BCNF
  – Decomposition into BCNF
3. SQL including Views

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)

Views
- Definition and how to use them
- Virtual v.s. materialized views
- View update problem

Know the syntax
Know the semantics (nested loops !)
4. Transactions

Transactions concepts
• What are the benefits and challenges of transactions?
  – Includes read-write and other conflicts
• What does ACID stand for
• 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
• Also: Protecting data beyond transactions (backups, etc.)
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
7. Data Integration and Cleaning

- Motivation for data integration
- Challenges of data integration
- Data integration approaches
- Problems with dirty data
- General approach to data cleaning
8. Relational Algebra and Datalog

• Relational algebra
  – Set semantics; make sure you review all joins!
  – Bag semantics: duplicate elimination, group-by
  – SQL to relational algebra

• Non-recursive datalog
  – What do multiple rules mean
  – What is a “safe” datalog rule
  – Translation to SQL
9. XML/XPath/XQuery

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