

CSE 344: Lecture 29

Final Review

The Final

- Wednesday, June 8, 8:30-10:30
MGH 231 (this room)
- Closed book
- Special office hours: Monday June 6, 10-11am,
CSE 303.

The Final

One problem for each week of classes

1. Basic SQL
2. Advanced SQL
3. Relational algebra/calculus
4. XML/XPath/XQuery
5. Conceptual Design, Constraints, Views
6. Transactions
7. Parallel data processing (parallel DB, MR, Pig)
8. Finding similar items

This is only an *approximate* match to the eight weeks of instructions

1. Basic SQL

- SELECT-FROM-WHERE
- DISTINCT, ORDER BY, renaming of attributes
- INSERT, DELETE, UPDATE

- Know the syntax
- Know the semantics (nested loops !)

2. Advanced SQL

- GROUP-BY
- HAVING: *different* from WHERE (why ?)
- NULLs, outer joins
- Nested queries (subqueries):
 - In the SELECT clause (e.g. aggregates)
 - In the FROM clause (temporary tables)
 - In the WHERE clause: EXISTS, NOT EXISTS

3. Relational Algebra/Calculus

- Relational algebra
 - Set semantics; make sure you review all joins !
 - Bag semantics: duplicate elimination, group-by
- Non-recursive datalog
 - What do multiple rules mean
 - What is a “safe” datalog rule
 - Translation to SQL
- Relational calculus
 - Existential, universal quantifier
 - Translation to non-recursive datalog (then to SQL)

4. XML/XPath/XQuery

- XML
 - Basic definitions: tags/elements/attributes/text, well-formed/valid XML document
 - Mapping between the relational model and back to XML (there are many ways to map !)
 - DTDs
- XPath – really easy...
- XQuery – a kind of SQL (seen one, seen all)

5. Conceptual Design, Constraints, Views

- E/R Diagrams:
 - Entities, attributes
 - Relationships:
 - Many-many, many-one, one-one
 - Multi-way relationships
 - Inheritance, weak entity sets, constraints
 - Translation to relations
 - Constraints in SQL
- Normal forms and functional dependencies:
 - Anomalies
 - Functional dependencies, X+
 - BCNF
- Views: definition; virtual v.s. materialized views

6. Transactions

- What does ACID stand for
- Concurrency control, Atomicity
- Definition of *serializability*
- The four isolation levels in SQL
- Locks v.s. MVCC
- Phantoms
- How to write serializable transactions in an MVCC system (like postgres, oracle)

7. Parallel Data Processing

- Parallel databases:
 - Speedup/scaleup
 - Shared memory, shared disk, shared nothing
 - How to implement simple algorithms: group-by, join
- Map/reduce
 - 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 Latin

8. Finding Similar Items

- The entity-resolution problem
- How to check if two items are similar:
 - Edit distance
 - Jaccard similarity
 - Size of the intersection, or symmetric difference
- Shingles, or q-grams
- Min-hashes
- LSH

General Advice

How to study:

- Go over the lecture notes, and try to remember what we discussed in class
- Go over the material in the sections
- Go over the homeworks
- Read the material in the book
- Review the midterm

Make sure you understand ! In doubt ? Come to see me Monday 10am-11am.