

# CS 4604: Introduction to Database Management Systems

*B. Aditya Prakash*

Midterm Review

# MIDTERM REVIEW

# Course Outline

- Weeks 1–4: Query/ Manipulation Languages and Data Modeling
  - Relational Algebra
  - Data definition
  - Programming with SQL
  - Entity-Relationship (E/R) approach
  - Specifying Constraints
  - Good E/R design
- Weeks 5–8: Indexes, Processing and Optimization
  - Storing
  - Hashing/Sorting
  - Query Optimization
  - NoSQL and Hadoop
- Week 9-10: Relational Design
  - Functional Dependencies
  - Normalization to avoid redundancy
- Week 11-12: Concurrency Control
  - Transactions
  - Logging and Recovery
- Week 13–14: Students' choice
  - Practice Problems
  - XML
  - Data mining and warehousing

# Course Outline: For Midterm Exam

- Weeks 1–4: Query/ Manipulation Languages and Data Modeling
  - Relational Algebra
  - Data definition
  - Programming with SQL
  - Entity-Relationship (E/R) approach
  - Specifying Constraints
  - Good E/R design
- Weeks 5–8: Indexes, Processing and Optimization
  - Storing
  - Hashing/Sorting
  - Query Optimization
  - NoSQL and Hadoop
- Week 9-10: Relational Design
  - Functional Dependencies
  - Normalization to avoid redundancy
- Week 11-12: Concurrency Control
  - Transactions
  - Logging and Recovery
- Week 13–14: Students' choice
  - Practice Problems
  - XML
  - Data mining and warehousing

No Query  
Processing/  
Optimization

# FUNDAMENTAL

## Relational operators

- selection

$$\sigma_{condition} (R)$$

- projection

$$\pi_{att-list} (R)$$

- cartesian product

$$R \times S$$

- set union

$$R \cup S$$

- set difference

$$R - S$$

# Relational ops

- Surprisingly, they are enough!
- Derived/convenience operators:
  - set intersection  $\cap$
  - **join** (theta join, natural join)  $\bowtie$
  - ‘rename’ operator  $\rho_{R'}(R)$
  - division  $R \div S$

# Basic SQL Query

```
SELECT [DISTINCT] target-list  
FROM relation-list  
WHERE qualification;
```

- **Relation-list:** A list of relation names (possibly with **range-variable** after each name).
- **Target-list:** A list of attributes of relations in relation-list
- **Qualification:** conditions on attributes
- **DISTINCT:** optional keyword for duplicate removal.
  - Default = no duplicate removal!
- **ORDER BY:** for sorting values

# SQL

- Handling Sub-queries
- SQL Data Definition Commands
- Constraints
- Triggers
- ...
- Note how referential integrity can be enforced (foreign key; on delete cascade etc.)



# E/R Diagrams

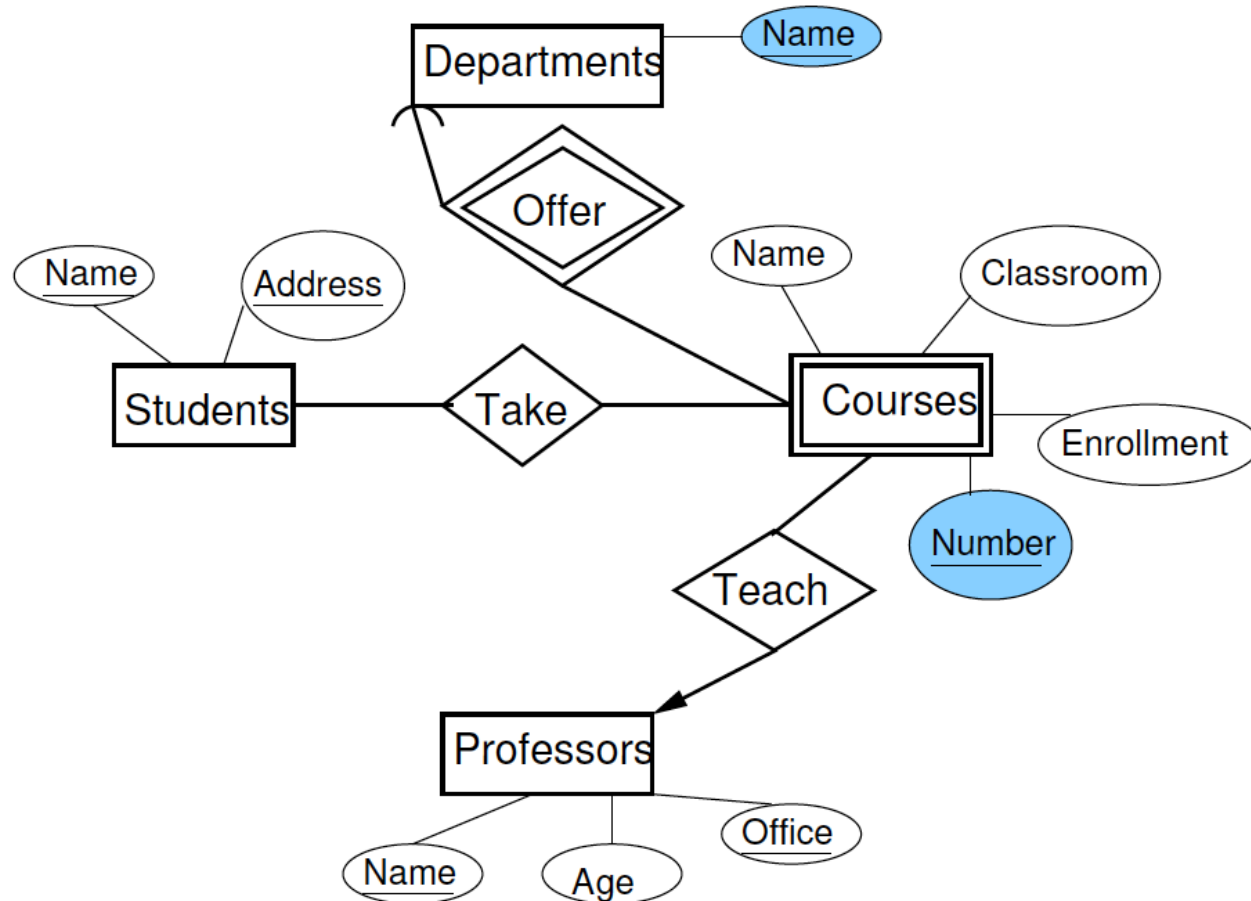
- IMPORTANT:
  - **Follow only lecture slides for this topic!**
  - Differences from the book:
    - More details
    - Slightly different notation

# Relationships

- Show a many-one relationship by **an arrow entering the “one” side.**      Many  $\longrightarrow$  One
- Show a one-one relationship by **arrows entering both entity sets.**      One  $\longleftrightarrow$  One
- In some situations, we can also assert **“exactly one,”** i.e., each entity of one set must be related to exactly one entity of the other set. To do so, we use a **rounded arrow.**       $\longrightarrow$  Exactly One

# E/R Example (does not contain ISA)

- Each department teaches multiple courses. Each course has a number. What is the key for the entity set Courses?



# Converting E/R Diagrams to Relational Designs

- Entity Set  $\rightarrow$  Relation
  - Attribute of Entity Set  $\rightarrow$  Attribute of a Relation
- Relationship  $\rightarrow$  relation whose attributes are
  - Attribute of the relationship itself
  - Key attributes of the connected entity sets
- Several special cases:
  - Weak entity sets.
  - Combining relations (especially for many-one relationships)
  - *ISA* relationships and subclasses
- Also note how referential integrity comes in (foreign keys)

# Tree Indexes

- B+-Trees
  - **Carefully understand the Definition!**
  - Searching
  - Inserting
  - Deleting

# Hashing/Sorting

- Extendible Hashing
  - Linear Hashing
  - External Sorting
- 
- Again, how to search and build, internalize the structure
  - Sorting: understand the process, how to cost it, how many passes it takes etc.

# Exam

- **No** aids allowed EXCEPT:
  - **Only written (not typed)** 1 letter-size page (you may use **both** sides)
  - A calculator (NOT your smartphone)
- Duration: 75 mins, during class March 2, Wednesday
- More or less equal weightage to all the topics
  - Questions will be similar to the HWs, Handouts