Today’s Plan

Quiz 3
Exam
Entity-Relationship Diagrams

Terms

• monotonic queries
  • Do not lose tuples as more data arrives (think streaming queries)

• table function
  • returns set of tuples
  • used anywhere a table is used
Quiz 3: tough problems

• 2. \((\text{NULL} = 20) \text{ or } (10 = 10)\) and \((\text{NULL} = 10)\) is unknown
  • true

• 9. “select A, max(B) from R” mixing scalar and non-scalar (no groupby for A)

• 12. Are the following two queries equivalent? Why or Why not? Assume R.a is an integer attribute.
  1. select * from R where R.a > 1;
  2. (select * from R) except (select * from R where R.a <= 1);

  • Check what happens if all A are nulls

Quiz 3: tough problems

**Q15**

0.2 Points

For three relations R(A, B), S(B, C), T(C, D), write relational algebra expressions to generate the following relations:

1. Q1\((A, D)\) where R and S are joined on condition R.B > S.B, and S and T have a natural join.
2. Q2\((A, C)\) to find all \((A, C)\) pairs such that R.B = S.B, and S.C does not have a matching tuple in T.

In both cases, use only the basic relational operations.

EXPLANATION

\[ Q1 \leftarrow \pi_{A,D} \left( \sigma_{R.B>S.B} (R \times S) \times T \right) \]
\[ Q2 \leftarrow \pi_{A,C} \left( (R \times (S \text{ antijoin } T)) \right) \]
Exam

• Covers through Quiz 5
• Online (no class)
• Examples will be on the schedule next week
  • not great examples, as all different because of remote learning.

Entity-Relationship Model
Entity-Relationship Model

Two key concepts

- **Entities:**
  - An object that exists and is distinguishable from other objects
    - Examples: Bob Smith, BofA, CMSC424
  - Have *attributes* (people have names and addresses)
  - Form *entity sets* with other entities of the same type that share the same properties
    - Set of all people, set of all classes
  - Entity sets may overlap
    - Customers and Employees
Entity-Relationship Model

Two key concepts

» Relationships:
  • Relate 2 or more entities
    – E.g. Bob Smith has account at College Park Branch
  • Form relationship sets with other relationships of the same type that share the same properties
    – Customers have accounts at Branches
  • Can have attributes:
    – has account at may have an attribute start-date
  • Can involve more than 2 entities
    – Employee works at Branch at Job

Baby ER Diagrams (illustration only, do not use)

Rectangles: entity sets
Diamonds: relationship sets
Ellipses: attributes
Rest of the class

Details of the ER Model
   » How to represent various types of constraints/semantic information etc.

Design issues

A detailed example

Next: Relationship Cardinalities

We may know:
   • One customer can only open one account
   • OR
   • One customer can open multiple accounts

Representing this is important

Why ?
   » Better manipulation of data
      • If former, can store the account info in the customer table
   » Can enforce such a constraint
      • “Application logic will handle it” NOT GOOD
   » If not represented in conceptual model, domain knowledge can easily be lost
Mapping Cardinalities

Express the number of entities to which another entity can be associated via a relationship set

Most useful in describing binary relationship sets

N-ary relationships?
» More complicated
» Details in the book

Mapping Cardinalities

One-to-One

One-to-Many

Many-to-One

Many-to-Many
Next: Types of Attributes

Simple vs Composite
» Single value per attribute?
  • Are parts accessed separately?
  • E.g. accessing first and last names from name

Single-valued vs Multi-valued
» E.g. Phone numbers are multi-valued

Derived
» If date-of-birth is present, age can be derived
» Can help in avoiding redundancy, enforcing constraints etc...

Types of Attributes
**Note!**

The boxes and ellipses diagrams shown here are just throwaway notation for describing entity sets and relationship set in the abstract.

You will never use this notation yourself. Instead, use the E-R diagrams discussed later in this set of slides.

---

**Next: Keys**

Key = set of attributes that uniquely identifies an entity or a relationship
**Entity Keys**

Possible Keys:
- \{cust-id\}
- \{cust-name, cust-city, cust-street\}
- \{cust-id, age\}

cust-name ?? Probably not.

Domain knowledge dependent !!

---

**Entity Keys**

*Superkey*

» any attribute set that can distinguish entities

*Candidate key*

» a minimal superkey

• Can’t remove any attribute and preserve key-ness
  – \{cust-id, age\} not a candidate key
  – \{cust-name, cust-city, cust-street\} is
    » assuming cust-name is not unique, and?

*Primary key*

» Candidate key chosen as *the* key by DBA
» Underlined in the ER Diagram
**Entity Keys**

{\textit{cust-id}} is a natural primary key

Typically, SSN forms a good primary key

Try to use a candidate key that rarely changes

» e.g. something involving address not a great idea

**Relationship Set Keys**

What attributes are needed to represent a relationship completely and uniquely?

» Union of primary entities keys, and relationship attributes

» \{\textit{cust-id}, access-date, account number\} describes a relationship completely
Relationship Set Keys

Is \{\text{cust-id, access-date, account number}\} a candidate key for \text{has}?

» No. Attribute \text{access-date} can be removed from this set without losing key-ness

» In fact, \text{union of primary keys of associated entities is always a superkey}

Almost always.

Relationship Set Keys

Is \{\text{cust-id, account-number}\} a candidate key?

» Depends...
Is \{\text{cust-id, account-number}\} a candidate key?

» Depends...

- If one-to-many relationship, \{\text{account-number}\} is a candidate key
  - Implies a customer can have many accounts, but at most one account holder per account.

- If one-to-one relationship, either \{\text{cust-id}\} or \{\text{account-number}\} sufficient
  - Since a given customer can only have one account, she can only participate in one relationship
  - Same for account
**Relationship Set Keys**

General rule for binary relationships
- one-to-one: primary key of either entity set
- one-to-many: primary key of the entity set on the many side
- many-to-many: union of primary keys of the associate entity sets

n-ary relationships
- More complicated rules

---

**Summary: Keys for Relationship Sets**

- Primary keys of entity sets together form super key of a relationship set.
  - \((s\_id, i\_id)\) is the super key of advisor
  - **NOTE:** this means a entity pair have at most one relationship in a relationship set.
    - Example: if we wish to track multiple meeting dates between a student and her advisor, we cannot assume a relationship for each meeting.
    - Fix: use a multivalued attribute.
- Must consider the mapping cardinality of the relationship set when identifying candidate keys.
- Consider relationship set semantics in selecting the primary key if more than one candidate key.
Entity-Relationship (E-R) Diagrams

Outline

Relational Algebra (6.1)
E/R Model (7.2 - 7.4)
E/R Diagrams (7.5)
Reduction to Schema (7.6)
Relational Database Design (7.7)
Functional Dependencies (8.1 – 8.4)
Normalization (8.5 – 8.7)
Relational Query Languages
SQL Basics
Formal Semantics of SQL
E-R Diagrams

- Rectangles represent entity sets.
- Diamonds represent relationship sets.
- Attributes listed inside entity rectangle
- Underline indicates primary key attributes

Composite, Multivalued, and Derived Attributes

- Composite, multivalued, and derived attributes are indicated in the diagram.
- Examples of composite attributes include:
  - instructor ID
  - name
  - first_name
  - middle_initial
  - last_name
  - address
  - street
  - street_number
  - street_name
  - apt_number
  - city
  - state
  - zip
  - phone_number
  - date_of_birth
  - age ( )
Relationship Sets with Attributes

Roles

- Entity sets of a relationship need not be distinct
  - Each occurrence of an entity set plays a “role” in the relationship
- The labels “course_id” and “prereq_id” are called roles.
Cardinality Constraints

- We express cardinality constraints by drawing either a directed line (→), signifying “one,” or an undirected line (—), signifying “many,” between the relationship set and the entity set.
- One-to-one relationship:
  - A student is associated with at most one instructor via the relationship advisor
  - A student is associated with at most one department via stud_dept

One-to-One Relationship

- one-to-one relationship between an instructor and a student
  - an instructor is associated with at most one student via advisor
  - and a student is associated with at most one instructor via advisor
**One-to-Many Relationship**

- one-to-many relationship between an *instructor* and a *student*
  - an instructor is associated with any number (including 0) of students via *advisor*
  - a student is associated with at most one instructor via advisor,

![Diagram](image1)

**Many-to-One Relationships**

- In a many-to-one relationship between an *instructor* and a *student*,
  - an instructor is associated with at most one student via *advisor*,
  - and a student is associated with several (including 0) instructors via *advisor*

![Diagram](image2)
Many-to-Many Relationship

- An instructor is associated with any number (possibly 0) of students via advisor
- A student is associated with any number (possibly 0) of instructors via advisor

Participation of an Entity Set in a Relationship Set

- Total participation (indicated by double line): every entity in the entity set participates in at least one relationship in the relationship set
- E.g., participation of section in sec_course is total
  - every section must have an associated course
- Partial participation: some entities may not participate in any relationship in the relationship set
- Example: participation of course in section is partial
Alternative Notation for Cardinality Limits

- Cardinality limits can also express participation constraints

```
<table>
<thead>
<tr>
<th>instructor</th>
<th>0..*</th>
<th>advisor</th>
<th>1..1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>salary</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>student</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tot_cred</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

E-R Diagram with a Ternary Relationship

```
instructor
| ID | name | salary |

student
| ID | name | tot_cred |

project

proj_guide

| ID | name | tot_cred |
```
Cardinality Constraints on Ternary Relationship

- We allow at most one arrow out of a ternary (or greater degree) relationship to indicate a cardinality constraint.
- E.g., an arrow from proj_guide to instructor indicates each student has at most one guide for a project.
- If there is more than one arrow, there are two ways of defining the meaning.
  - E.g., a ternary relationship $R$ between $A$, $B$ and $C$ with arrows to $B$ and $C$ could mean
    1. each $A$ entity is associated with a unique entity from $B$ and $C$ or
    2. each pair of entities from $(A, B)$ is associated with a unique $C$ entity, and each pair $(A, C)$ is associated with a unique $B$
  - Each alternative has been used in different formalisms.
  - To avoid confusion we outlaw more than one arrow.

Weak Entity Sets

- An entity set that does not have a primary key is referred to as a weak entity set.
- The existence of a weak entity set depends on the existence of an identifying entity set.
  - It must relate to the identifying entity set via a one-to-many relationship set from the identifying to the weak entity set.
  - Weak side of relationship set must be total.
  - Identifying relationship depicted using a double diamond.
- The discriminator (or partial key) of a weak entity set is the set of attributes that help distinguish among all the entities of a weak entity set.
- The primary key of a weak entity set is formed by the primary key of the strong entity set on which the weak entity set is existent-dependent, plus the weak entity set’s discriminator.
Weak Entity Sets (Cont.)

• We underline the discriminator of a weak entity set with a dashed line.

• We put the identifying relationship of a weak entity in a double diamond.

• Primary key for section – (course_id, sec_id, semester, year)

Weak Entity Sets (Cont.)

• Note: the primary key of the strong entity set is not explicitly stored with the weak entity set, since it is implicit in the identifying relationship.

• If course_id were explicitly stored, section could be made a strong entity, but then the relationship between section and course would be duplicated by an implicit relationship defined by the attribute course_id common to course and section
Summary of E-R Notation

- **E** entity set
  - A1
  - A2
  - A2.1
  - A2.2
  - [A3]
  - A4

- **R** relationship set
  - identifying relationship set for weak entity set
    - primary key
  - total participation of entity set in relationship
    - discriminating attribute of weak entity set

- **relations:** attributes: simple (A1), composite (A2) and multivalued (A3) derived (A4)
Redundant Attributes

• Suppose we have entity sets
  - instructor, with attributes including dept_name
  - department
  and a relationship
  - inst_dept relating instructor and department

• Attribute dept_name in entity instructor is redundant since there is an explicit relationship inst_dept which relates instructors to departments
  - The attribute replicates information present in the relationship, and should be removed from instructor
  - BUT: redundant attributes sometimes get reintroduced when converting back to tables
Reduction to Relational Schemas

Outline

Relational Algebra (6.1)
E/R Model (7.2 - 7.4)
E/R Diagrams (7.5)
Reduction to Schema (7.6)
Relational Database Design (7.7)
Functional Dependencies (8.1 – 8.4)
Normalization (8.5 – 8.7)
Reduction to Relation Schemas

- Entity sets and relationship sets can be expressed uniformly as *relation schemas* that represent the contents of the database.
- A database which conforms to an E-R diagram can be represented by a collection of schemas.
- For each entity set and relationship set there is a unique schema that is assigned the name of the corresponding entity set or relationship set.
- Each schema has a number of columns (generally corresponding to attributes), which have unique names.

Representing Entity Sets With Simple Attributes

- A strong entity set reduces to a schema with the same attributes `student(ID, name, tot_cred)`
- A weak entity set becomes a table that includes a foreign key for the primary key of the identifying strong entity set `section (course_id, sec_id, sem, year)`
Representing Relationship Sets

• A many-to-many relationship set is represented as a schema with attributes for the primary keys of the two participating entity sets, and any descriptive attributes of the relationship set.

• Example: schema for relationship set advisor

\[ \text{advisor} = (s_{id}, i_{id}) \]