CTEC323 Lecture 4

Dan Zingaro
OISE/UT

September 25, 2008

Relational Model

- Relational model has three well-defined components
  - Logical data structure (tables)
  - Integrity rules for enforcing consistency
  - Operations that define how data are manipulated
- In the relational data model, we focus on logical representation instead of physical storage details

Tables

- From a conceptual point of view, tables resemble files
- A table contains a group of entities (an entity set)
- Rows represent single entity occurrences
- Columns represent attributes
- Each column has a specific range of values known as the attribute domain.
- Order of rows and columns is immaterial
- A table must have one or more attributes that uniquely identify each row

Keys

- A key consists of one or more attributes that determine other attributes
- A determines B means that given A, you can look up the unique value of B (e.g. student number determines last name)
- B is functionally dependent on A if A determines B
- Multi-attribute keys are called composite keys
- Any attribute that is part of a key is a key attribute
- Full functional dependence: B is functionally dependent on a composite key A but not on any subset of A

Exercise 1: keys

Imagine we are maintaining a catalog for a library. Our records have the following fields (the ones in bold are unique for each book).

Call Number	Title	Publication Date	Author	**ISBN**

- Which attributes does Call Number determine?
- Which attributes does Author determine?
- Which attributes does the combination of Call Number and Author determine?
- Is Publication Date fully functionally dependent on the combination of Call Number and Author?

Keys...

- Super key: uniquely identifies each row
- Candidate key is a minimal super key
- Primary key is the chosen candidate key designated to be the row identifier
- Entity integrity requires that each value of primary key is unique and not null
- A foreign key is an attribute whose values match the primary key in a related table
- Referential integrity requires that if foreign key contains a value, it refers to an existing row in another relation

Exercise 2: keys

Call Number Title Publication Date Author **ISBN**

- What are the super keys?
- What are the candidate keys?
- What are our choices for the primary key?

Relational Algebra

- Tables are manipulated through relational algebra
- Use of the following relational operators on relations yields new relations
- Union combines all rows from two tables; the tables must be union-compatible (same column names and types)
- Intersect yields only rows appearing in both tables; tables must be union-compatible
- Difference yields all rows that are in the first table but not the second; tables must be union-compatible
- Product (or Cartesian product) yields all possible pairs of rows from two tables
- Select or restrict (confusingly, “where” in SQL!) yields a horizontal subset of the table (i.e. include only rows meeting criteria)
- Project (“select” in SQL) yields a vertical subset of the table (all values for selected attributes)

Exercise 1: Relational Algebra

T1	Employee ID	Name
1	Dan	
2	Steph	
3	Joe	

T2	Employee ID	Name
2	Steph	
1	Joe	
3	Dave	
4	Dan	

- What is T1 union T2? T1 intersect T2? T1 minus T2? T2 minus T1? T1 product T2?
- What is the result of restricting T1 on ID $<=1$?

Relational Algebra...

- Divide operator takes a two-column table x as its first argument and a one-column table y as its second argument
- x relates each a in the first column to a set of values b in the second column
- The result of the divide is a single column table consisting of those elements a where b contains at least all elements in y
- Example: x maps student numbers to types of fees owed. y is a list of all possible fee types
- Then, dividing x by y yields those unfortunate students who must pay each type of fee

Relational Algebra...

- Join allows information to be combined from two or more tables
- Natural join links tables by selecting only the rows with common values in their common attributes
- Three-step process
  - Take product of tables
  - Perform a select yielding those rows where the join columns match
  - Project to yield a single copy of each attribute (avoiding duplicate columns)

Relational Algebra...

- An equijoin links tables based on an equality condition that compares specified columns
- If any comparison besides equality is used, the join is a theta join
- The outcome of an equijoin or theta join does not eliminate duplicate columns
- In an inner join, only matched rows are retained (like in a natural join)
- In an outer join, unmatched pairs of rows are retained and unmatched values in the other table are set to null

Exercise 2: Relational Algebra

T1	Employee ID	Name
1	Dan	
2	Steph	
3	Joe	
4	Dave	

T2	Employee ID	Salary
1	30000	
2	20000	
3	100000	
5	85000	

- What results from a natural join? Inner join on equal Employee IDs? T1 outer join T2 on equal Employee IDs? T2 outer join T1 on equal Employee IDs? Full join on equal Employee IDs?

Relationships

- The 1:m relationship is the relational ideal
- The 1:1 relationship should be rare
- The M:M relationship cannot be directly implemented in the relational model
- 1:M is implemented by putting primary key of the "one" side in the "many" table as a foreign key

Relationships...

- A direct implementation of M:M would result in duplication and efficiency concerns
- Instead, use two 1:M relationships, and introduce a bridge entity, whose foreign key is at least the primary keys from the two related tables
- For a primary key, the combination of the tables' primary keys can be used, or a new one can be created
- The bridge table can also contain other attributes (e.g. grade for a student in a course)

Foreign Keys

- Proper use of foreign keys minimizes data redundancies
- Three often contradictory requirements
  - Design elegance
  - Processing speed
  - Information requirements
- Planned redundancy is often a part of good database design