

Lesson 4: Indexing and Relationships

Indexing

- A hugely important mechanism for optimizing the performance a RDBMS, especially for tables with many rows.
- Imagine trying to find a word in a book without an index, this is the same concept for a database.
- We'll dive deeper into indexing later on.

Attributes

Domains

- Recall that a domain is the range of all values for a given attribute.
 - The domain could be known and (more or less) static, such as for a GPA attribute.
 - A domain could also be dynamic and not necessarily known, such as attribute holding the Date of Birth of everyone at SUNY Orange.

Required vs Optional Attributes

- Hey, NULLs again!?
- **Required Attributes** are those that must be populated with a set of non-NULL values.
- **Optional Attributes** are those that could be populated with NULL values.

Identifiers

- Identifiers and composite identifiers are synonyms to key terms for the last two lessons:
 - A primary key is an attribute or combination of attributes that uniquely identifies any given row (entity instance).
 - Composite keys are made up of more than one attribute.

Simple vs. Composite Attributes

- **Composite Attributes** could be broken down into multiple other attributes.
- Whether these attributes should be broken down more should be decided during data modelling with the end users.

address
115 South Street, Middletown, NY 10940

address_num	street	city	state	zip_code
115	South Street	Middletown	NY	10940

Simple vs. Composite Attributes (continued)

- **Simple Attributes** are attributes that could not be broken down any further.

Single-valued Attributes

- Attributes that could have one and only one value for a single entity occurrence.
 - Example: A social security number is a good example of this.

Multi-valued Attributes

- Attributes that can have many values for a single entity occurrence.
 - Example: We might have several phone numbers.

Derived Attributes

- Can also be known as calculated or computed attributes.
- Example: An attribute like a person's age is a good candidate for a derived attribute. It's going to change once per year and it shouldn't be updated manually.
- Depending on the values and equation, these values could be stored (persisted) or generated (on the fly) when the attribute is referenced.
- We'll cover this more hands-on later!

Relationships

Cardinality vs. Connectivity (Continued)

Cardinality	Connectivity
Minimum and maximum number of entity occurrences	One-to-One
	One-to-Many
	Many-to-Many

Cardinality

- Ensuring the minimum and maximum values of cardinality is not handled at the table level in a DBMS.
- Applications or triggers (future topic!) need to preserve the range defined in the cardinality.

Existence Dependence

- **Existence dependence** occurs if an entity can exist in the database only when it is associated with another related entity occurrence.
- What does this mean?
 - Recall the concept of referential integrity.
 - Let's write out an example!

Existence Independence

- **Existence independence** is when an entity can exist apart from related entities.
- Entities that are existence independent are called **strong entities**.

Weak Relationships

- These exist if the primary key of a related entity doesn't contain a primary key component of the parent entity.
- They are also known as non-identifying relationships.
- Let's work through an example

Weak Entities

- They are existence dependent.
- They have a primary key that is partially or totally derived (inherited) from that of the parent entity of the relationship.

Relationship Participation

- Entities that participate in a relationship are known as participants.
- Similar concept to optional vs. required attributes, but instead used to govern the relationship.
- **Optional participation** means that one entity occurrence doesn't require corresponding entity occurrence in a particular relationship.
- **Mandatory participation** means that one entity occurrence requires a corresponding entity occurrence in a particular relationship.
- Let's look at an example!

Relationship Symbolology

Relationship Degree

- The **relationship degree** indicates the number of participants associated with a relationship

Relationship Degree	Number of Participants
Unary	1
Binary	2
Ternary	3
Higher Order	>3

Relational Schema

- Used for writing out the the relational schema.
- Example: table_name(primary_key_attribute, attribute1, attribute2,...,attributen)

Data Dictionaries

- Documentation, documentation, documentation!
- Recall the output of the data modeling process is a data model likely in the form of an ERD. This functions as only a portion of the documentation that we need.
- Data dictionaries provide detailed and human-readable descriptions of entities and attributes, identify constraints, business definitions and more!
- Doing a data dictionary as you go, is much easier than trying to put one together at the end.

Homework

- Read Chapter 5