
7

DATA MODELING AND ANALYSIS

Chapter Map



System Models

A model is a **representation of reality**. **HOW?**

Logical models show **what** a system is or does. They are implementation independent; that is, they depict the system independent of any technical implementation.

Physical models show not only what a system is or does, but also **how** the system is physically and technically implemented.

Data Modeling

Data modeling is a **technique for *organizing and documenting* a system's data**. Data modeling is sometimes called **database modeling** because a data model is eventually implemented as a database. It is sometimes called **information modeling**.

The **actual model** is frequently called an entity relationship diagram (**ERD**) because it depicts data in terms of the entities and relationships described by the data.

Sample Entity-Relationship Diagram (ERD)

CUSTOMER

Customer Number (Primary Key)
Customer Name
Shipping Address
Billing Address
Balance Due

ORDER

Order Number (Primary Key)
Order Date
Order Total Cost
Customer Number (Foreign Key)

ORDERED PRODUCT

Ordered Product ID (Primary Key)
.Order Number (Foreign Key)
.Product Number (Foreign Key)
Quantity Ordered
Unit Price at Time of Order

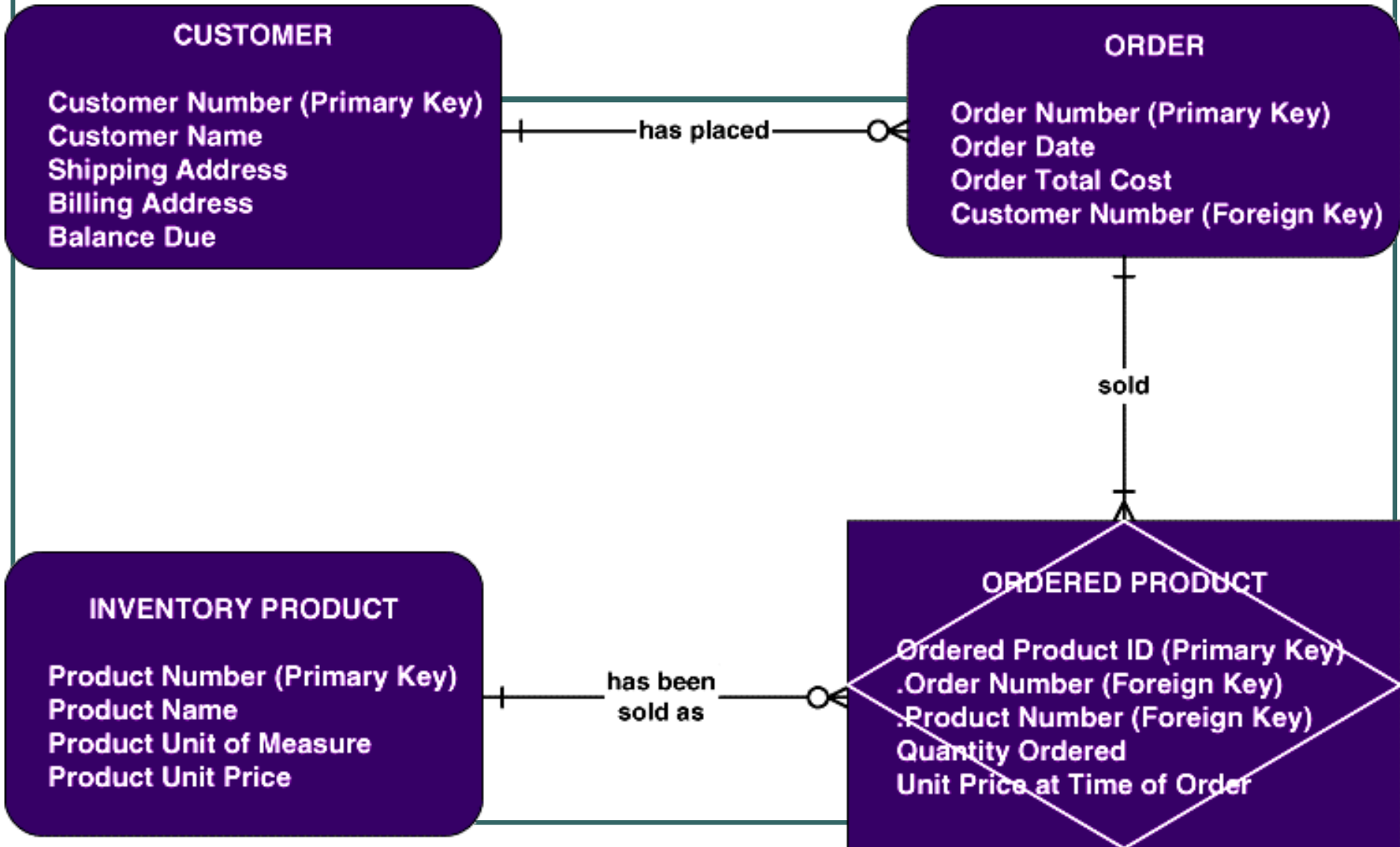
INVENTORY PRODUCT

Product Number (Primary Key)
Product Name
Product Unit of Measure
Product Unit Price

has placed

sold

has been
sold as



Data Modeling Concepts: Entity

An **entity** is a class of persons, places, objects, events, or concepts about which we need to capture and store data.

Name of Entity

- ☒ Persons: agency, contractor, customer, department, division, employee, instructor, student, supplier.
- ☒ Places: sales region, building, room, branch office, campus.
- ☒ Objects: book, machine, part, product, raw material, software license, software package, tool, vehicle model, vehicle.
- ☒ Events: application, award, cancellation, class, flight, invoice, order, registration, renewal, requisition, reservation, sale, trip.
- ☒ Concepts: account, block of time, bond, course, fund, qualification, stock.

DFD V.s. ERD

- **DFD** : a logical modeling tool
 - show the relationship between processes and data.
 - Process imply data, data imply processes.
- **ERD** : stresses the data and show the system's primary data entities are related.
 - quickly obtaining, with minimum efforts, a good sense of the structure of a database.

Data Modeling Concepts: Entity

An **entity instance** is a single occurrence of an entity.

Example: instances of the entity **STUDENT** may include

-  Betty Arnold
-  John Taylor
-  Lisa Simmons
-  Bill Macy
-  Heather Leath
-  Tim Wrench

What's the difference
between **entity** and
entity instance?

Data Modeling Concepts: Attributes

An **attribute** is a descriptive property or characteristic of an entity. Synonyms include **element**, **property**, and **field**.

A **compound attribute** is one that actually consists of other attributes

STUDENT

Name

.Last Name

.First Name

.Middle Initial

Address

.Street Address

.City

.State or Province

.Country

.Postal Code

Phone Number

.Area Code

.Exchange Number

.Number Within Exchange

Date of Birth

Gender

Race

Major

Grade Point Average

Data Modeling Concepts: Domains

The **data type** for an attribute defines what type of data can be stored in that attribute.

The **domain** of an attribute defines what **values** an attribute can legitimately take on.

The **default value** for an attribute is the value that will be recorded if not specified by the user.

Data Modeling Concepts: Identification

- 1 A **key** is an attribute, or a group of attributes, that assumes a **unique** value for each entity instance.
- 2 A **group of attributes** that uniquely identifies an instance of an entity is called a **concatenated key**.
- 3 A **candidate key** is a “candidate to become the primary key” of instances of an entity.
- 4 A **primary key** is that candidate key that will most commonly be used to uniquely identify a single entity instance.
- 5 Any candidate key that is not selected to become the primary key is called an **alternate key**.
- 6 A **subsetting criteria** is an attribute (or concatenated attribute) whose finite values divide all entity instances into useful subsets, e.g. **male Vs female**

Data Modeling Concepts: Identification Keys & Subsetting Criteria

STUDENT

Student Number (Primary Key)

Social Security Number (Alternate Key)

Name

.Last Name

.First Name

.Middle Initial

Address

.Street Address

.City

.State or Province

.Country

.Postal Code

Phone Number

.Area Code

.Exchange Number

.Number Within Exchange

Date of Birth

Gender (Subsetting Criteria 1)

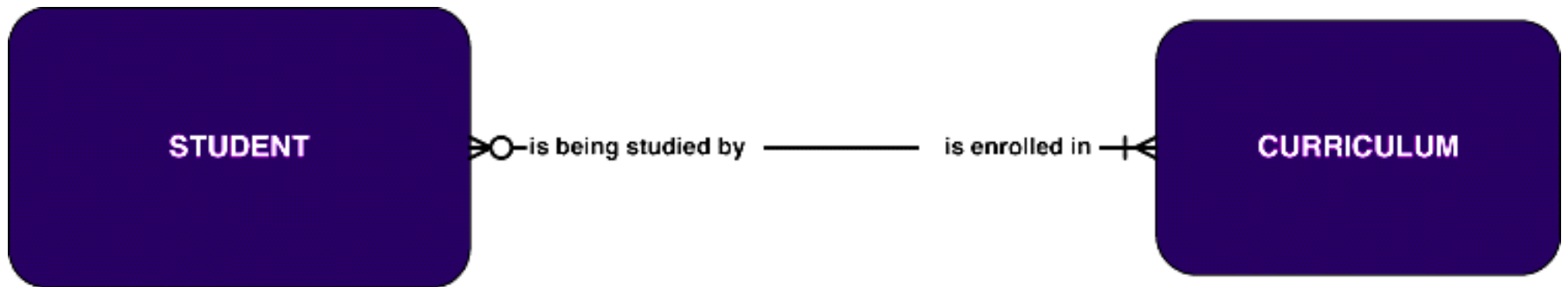
Race (Subsetting Criteria 2)

Major (Subsetting Criteria 3)

Grade Point Average

Data Modeling Concepts: Relationships

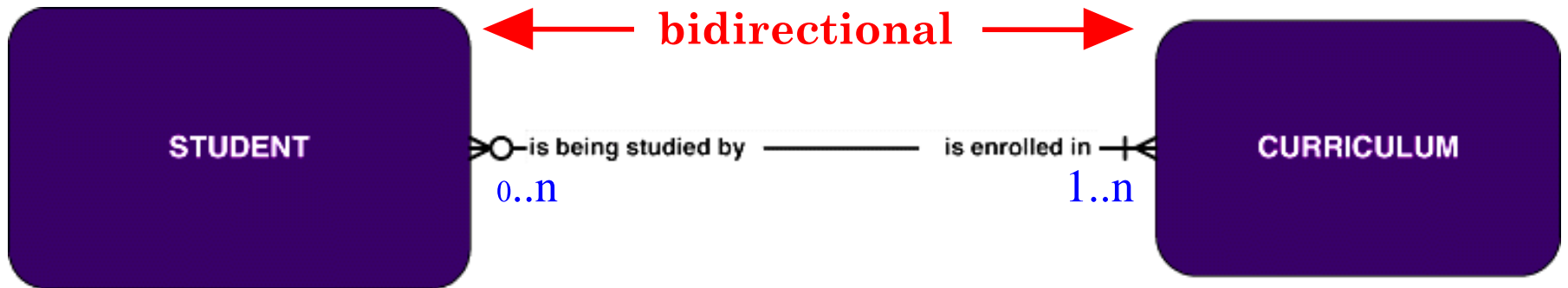
A **relationship** is a natural business **association** that exists between one or more entities. The relationship may represent an event that links the entities or merely a logical affinity that exists between the entities.



Data Modeling Concepts: Cardinality

Cardinality defines the minimum and maximum **number of occurrences** of one entity that may be related to a single occurrence of the other entity.

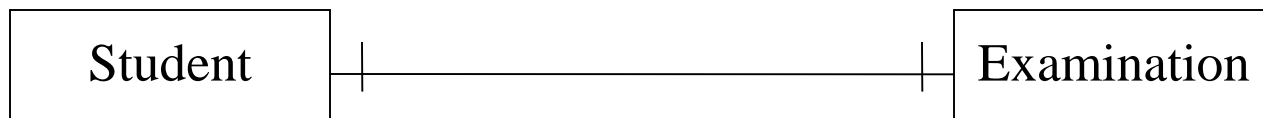
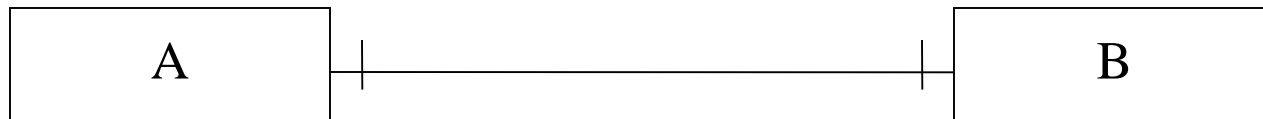
Because all relationships are bidirectional, cardinality must be defined in both directions for every relationship.



E.g. Write a sentence from the above diagram.

1:1 Relationship

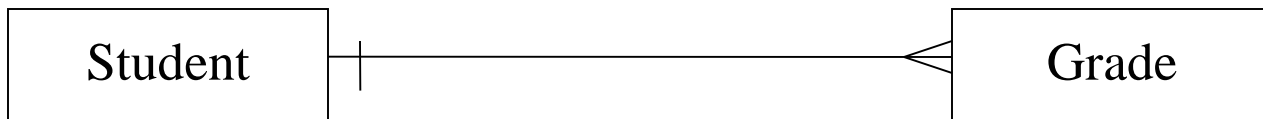
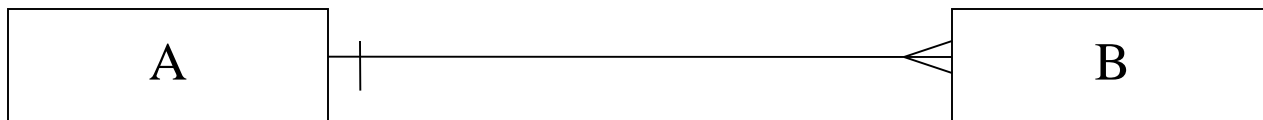
Each occurrence of entity A is associated with one and only one occurrence of entity B, and vice-versa.



Select only one question

1:m Relationship

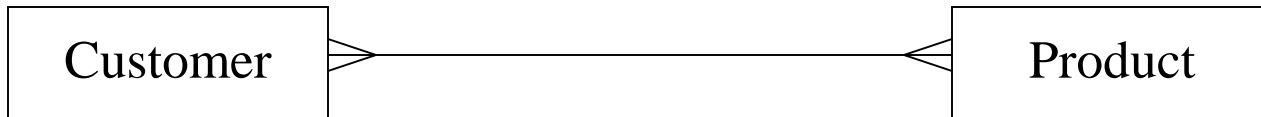
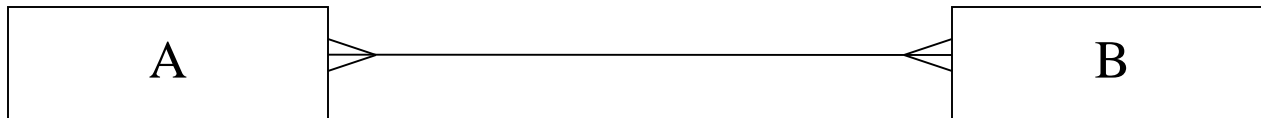
- Each occurrence of entity A is associated with one or more occurrence of entity B, but each occurrence of entity B is associated with only one occurrence of entity A.



course, paper, project

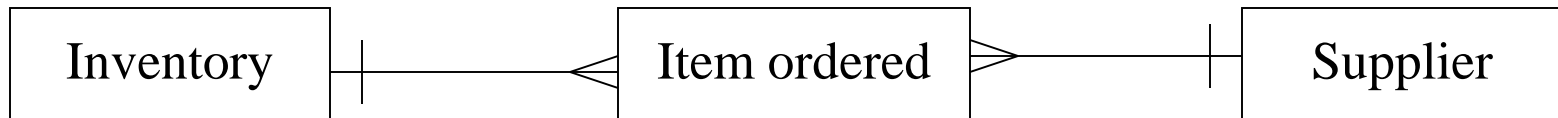
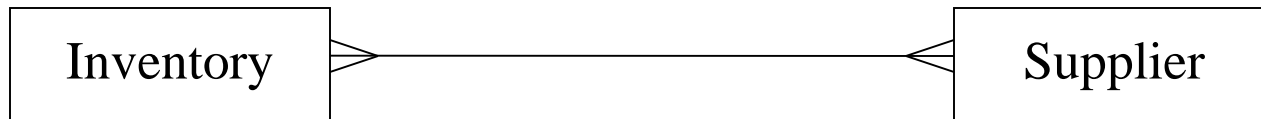
m:n Relationship

- Each occurrence of entity A is associated with one or more occurrence of entity B, and vice-versa.



Analyzing Relationship

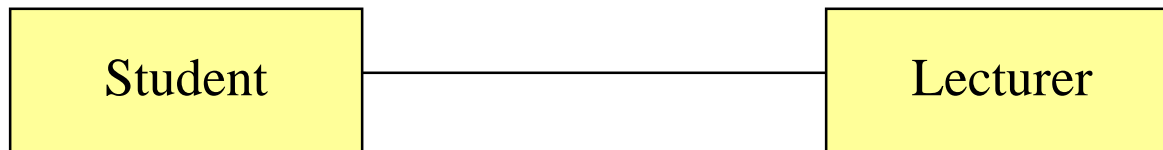
- m:n Relationship
- cause the maintenance problem.



Data Modeling Concepts: Degree

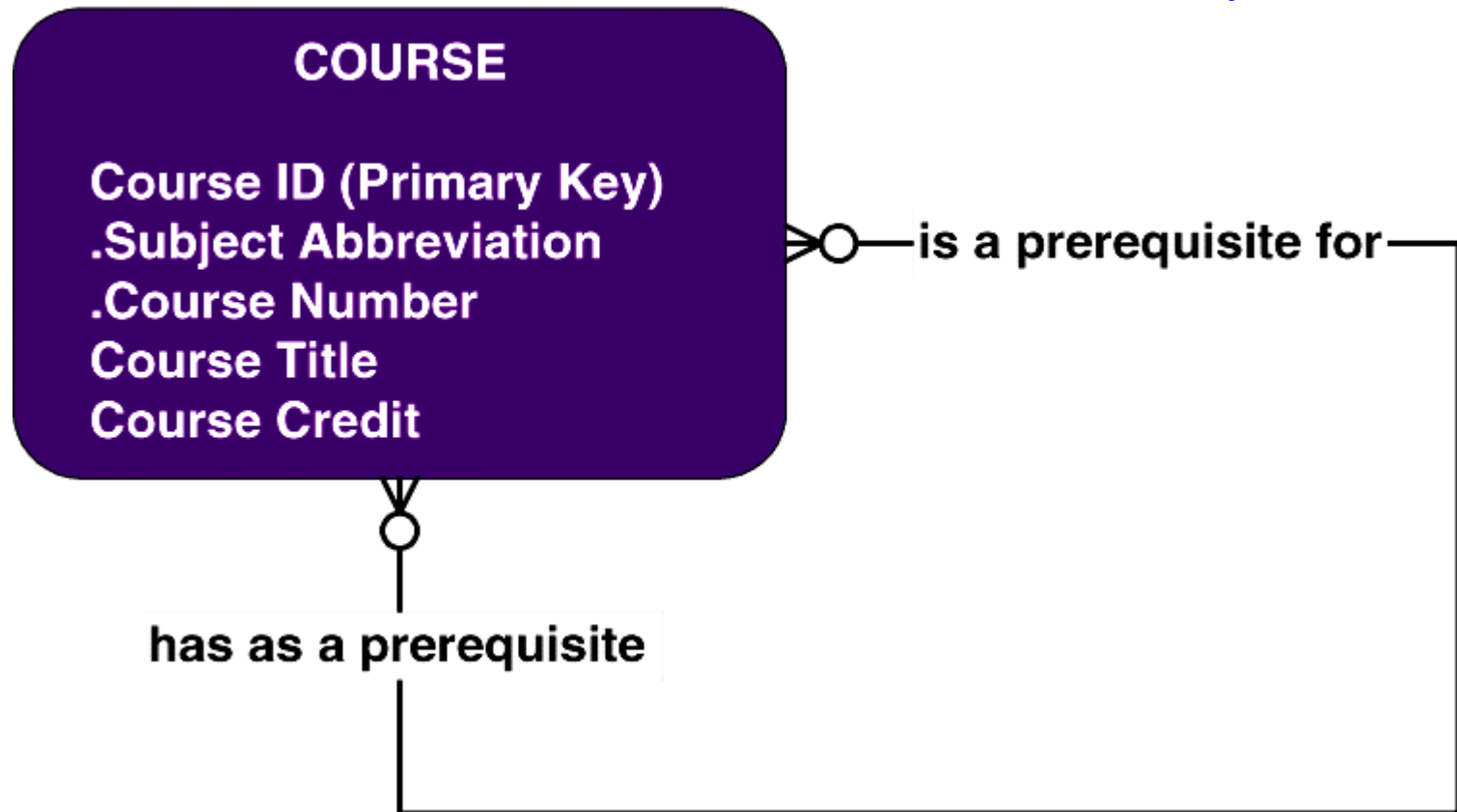
The **degree** of a relationship is the **number of entities** that participate in the relationship.

Binary



Data Modeling Concepts: Degree

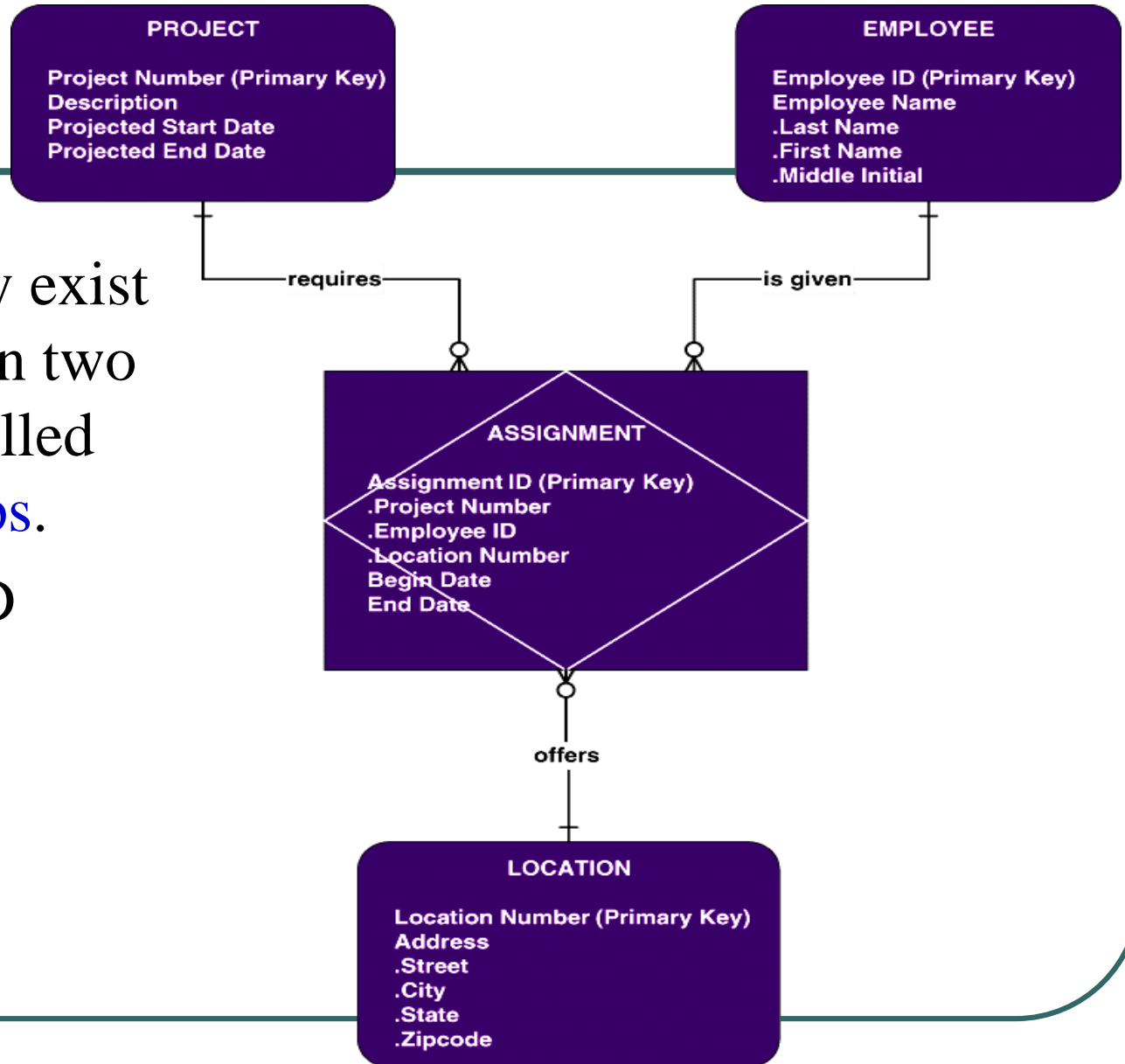
A **recursive** relationship is a relationship that exists between different instances of the **same entity**



Data Modeling Concepts: Degree

Relationships may exist between more than two entities and are called **N-ray relationships**.

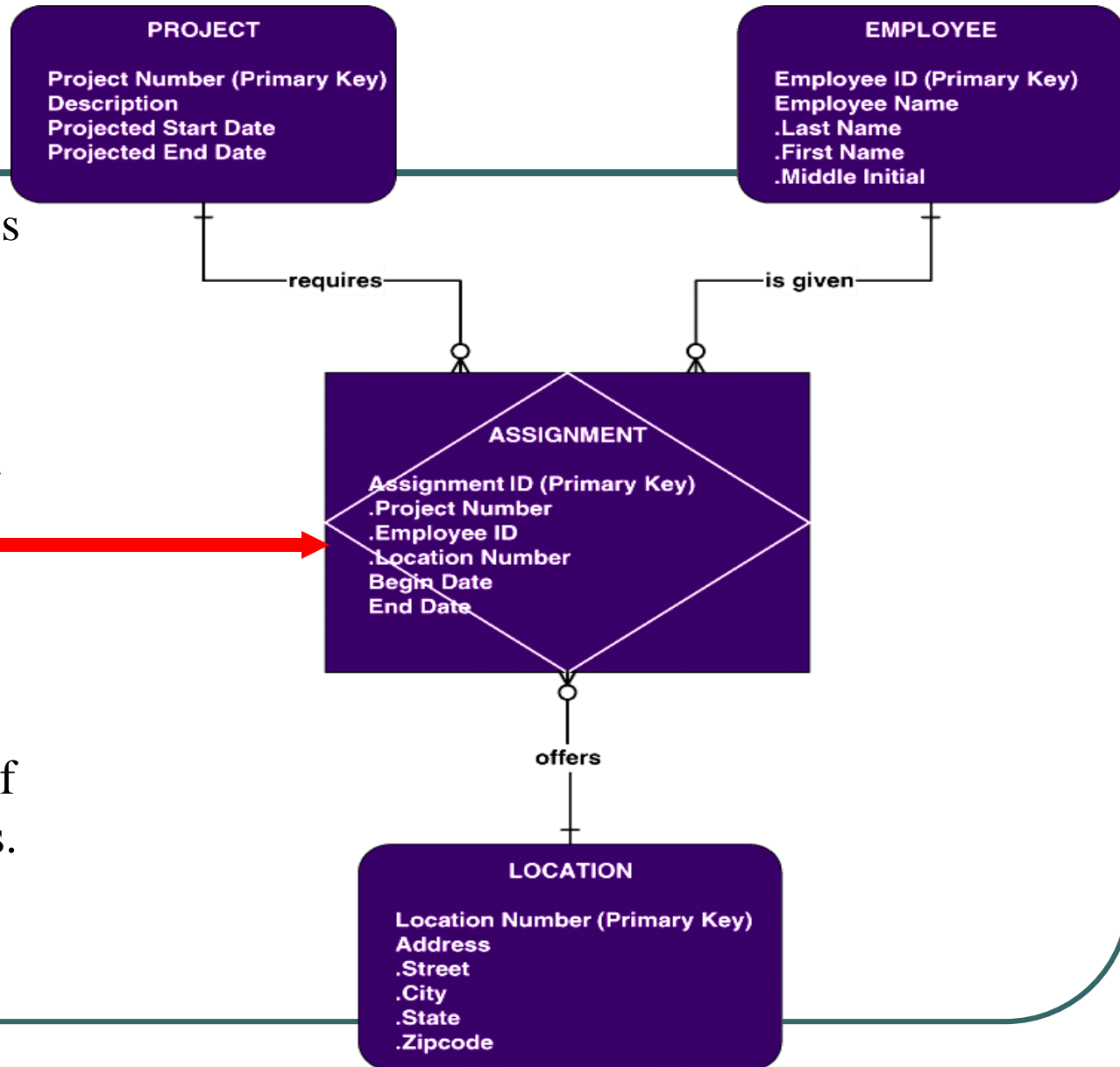
The example ERD depicts a ternary relationship.



Data Modeling Concepts: Degree

An **associative entity** is an entity that **inherits its primary key** from more than one other entity (called parents).

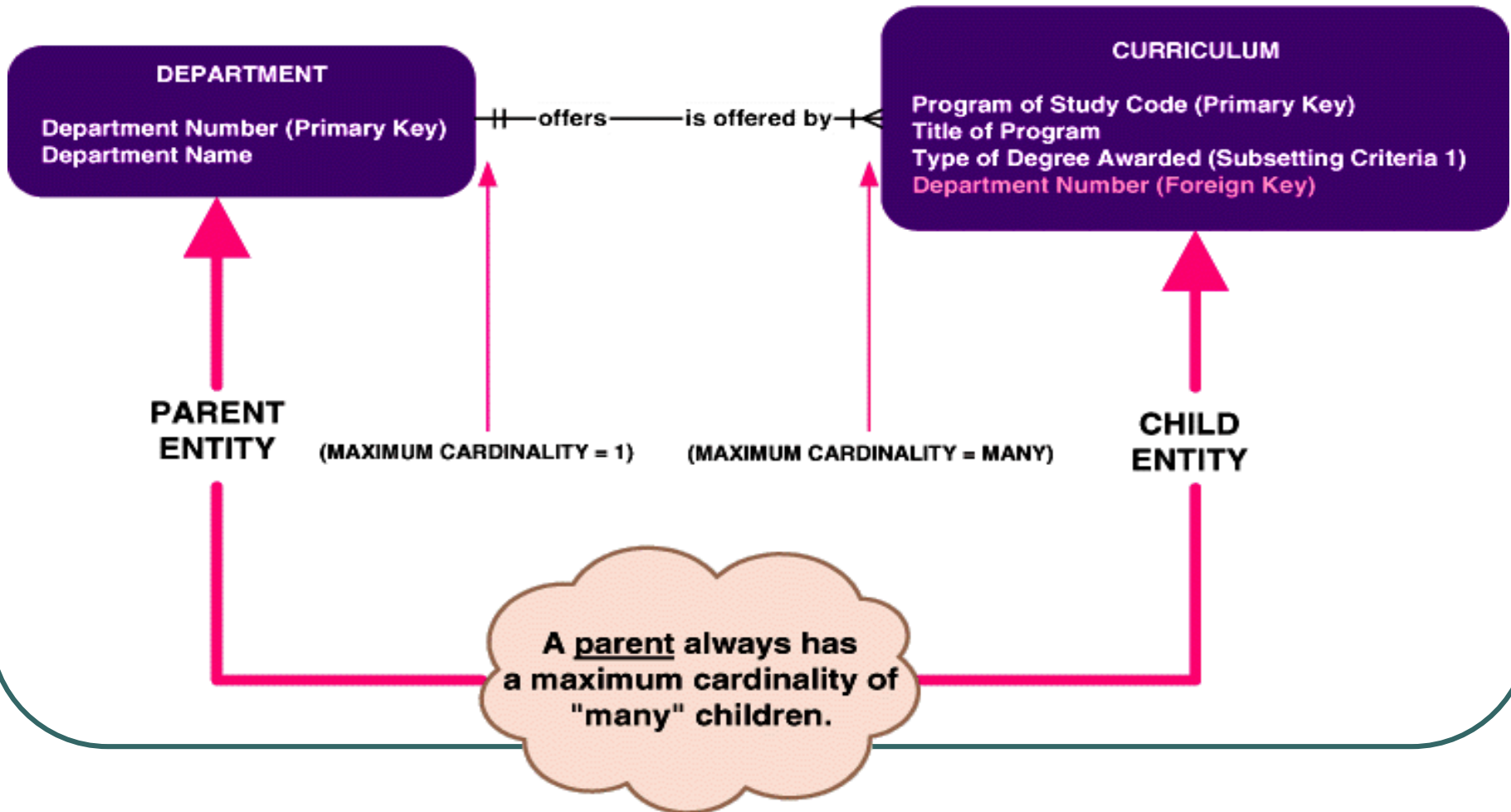
Each part of that concatenated key points to one and only one instance of each of the connecting entities.



Data Modeling Concepts: Foreign Keys

A **foreign key** is a primary key of one entity that is contributed to (duplicated in) another entity to identify instances of a relationship.

(a)

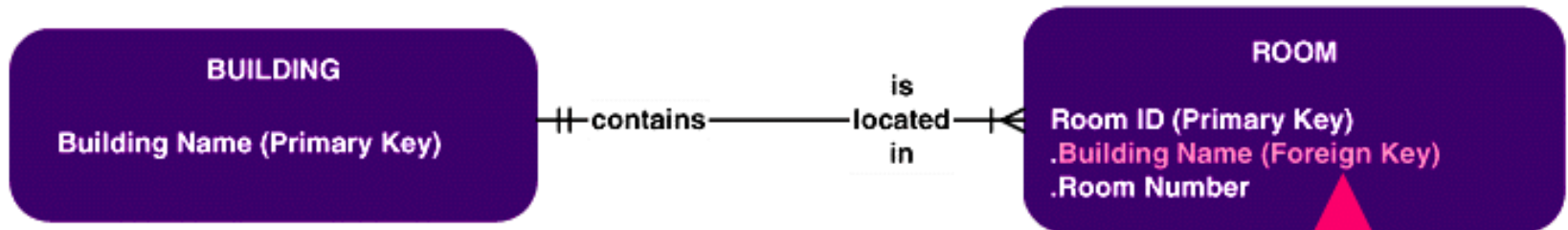


Data Modeling Concepts: Foreign Keys

Nonidentifying relationships are those in which each of the participating entities has its **own independent primary key**, In other words, none of the primary key attributes is shared.

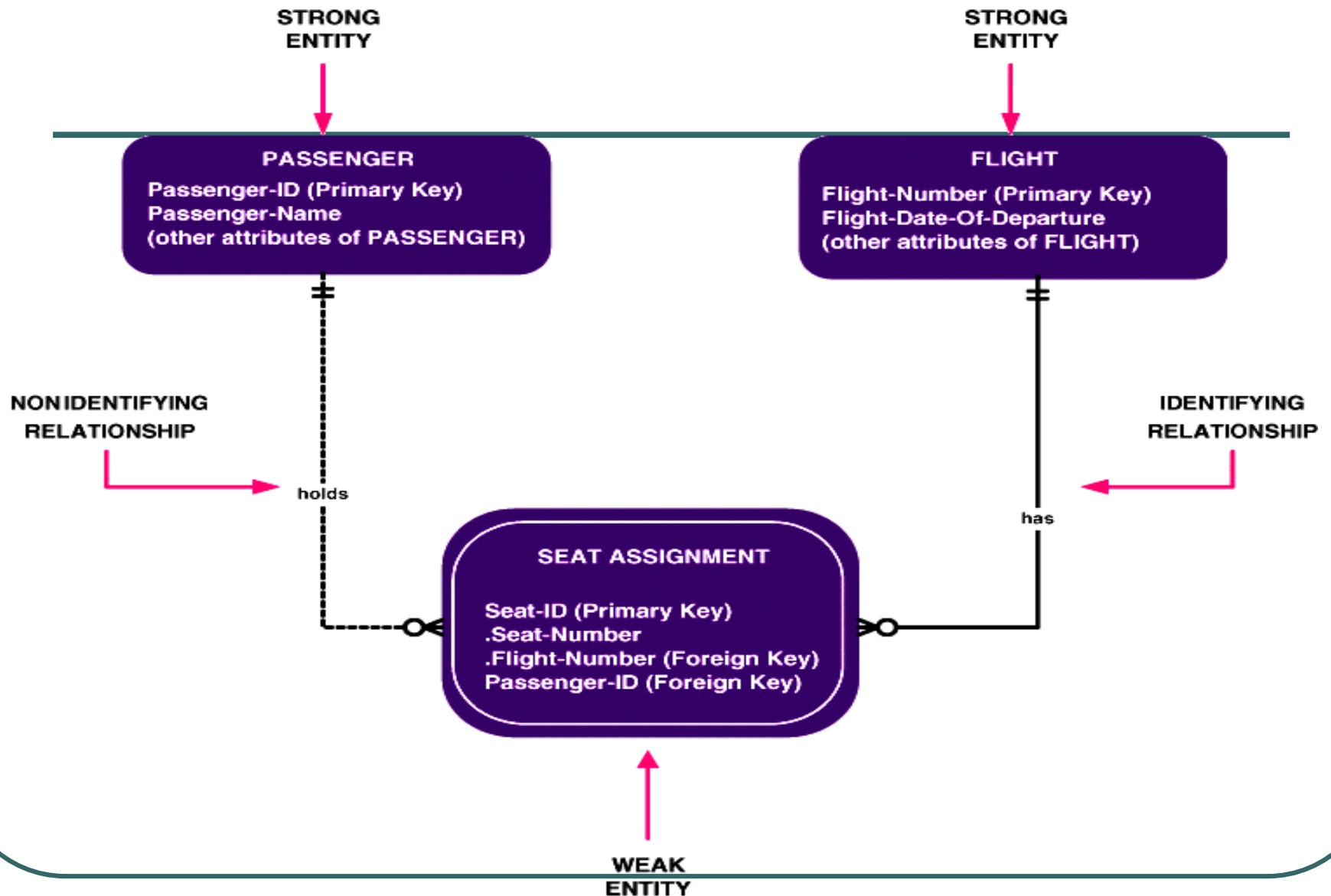
Identifying relationships are those in which the parent entity contributes its primary key to become **part of the primary key** of the child entity.

(b)



Identifying relationships have the foreign key participating as part of the child entity's *primary key*.

Data Modeling Concepts: Foreign Keys

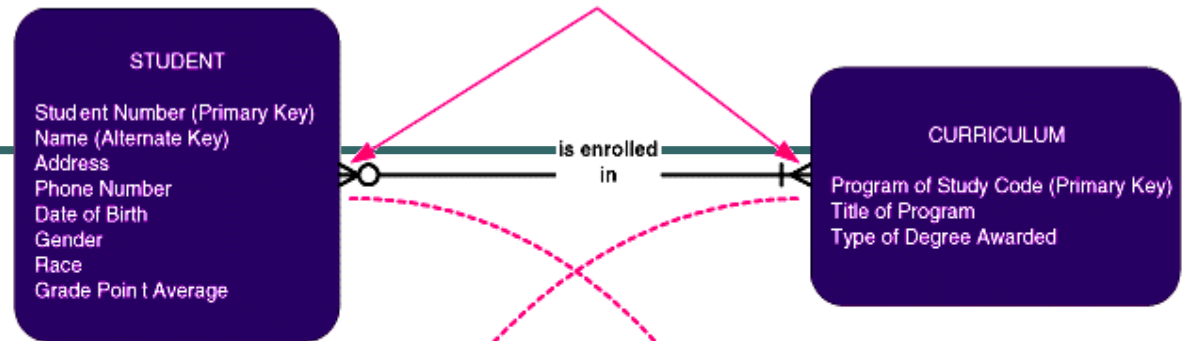


Data Modeling Concepts: Foreign Keys

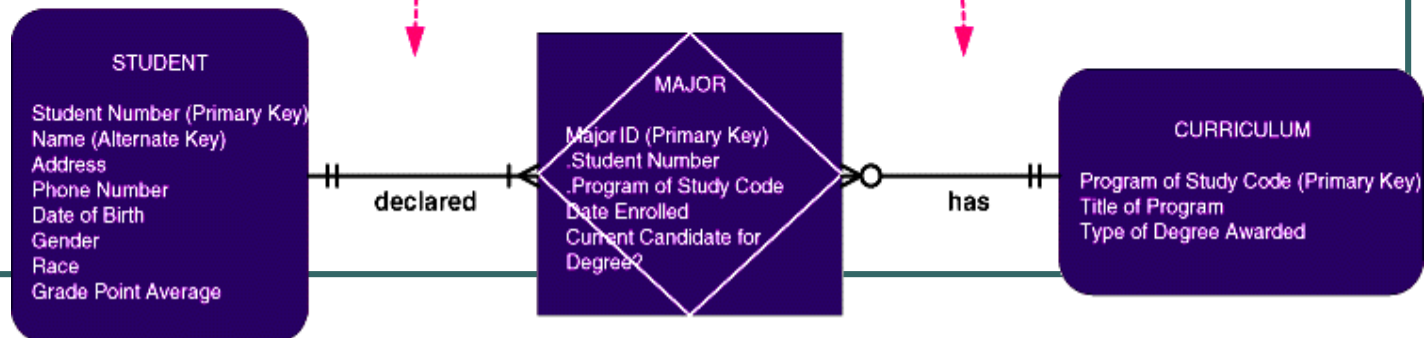
A **nonspecific relationship** (or **many-to-many relationship**) is one in which many instances of one entity are associated with many instances of another entity.

Nonspecific relationships must be resolved. Most nonspecific relationships can be resolved by introducing an associative entity.

(a)



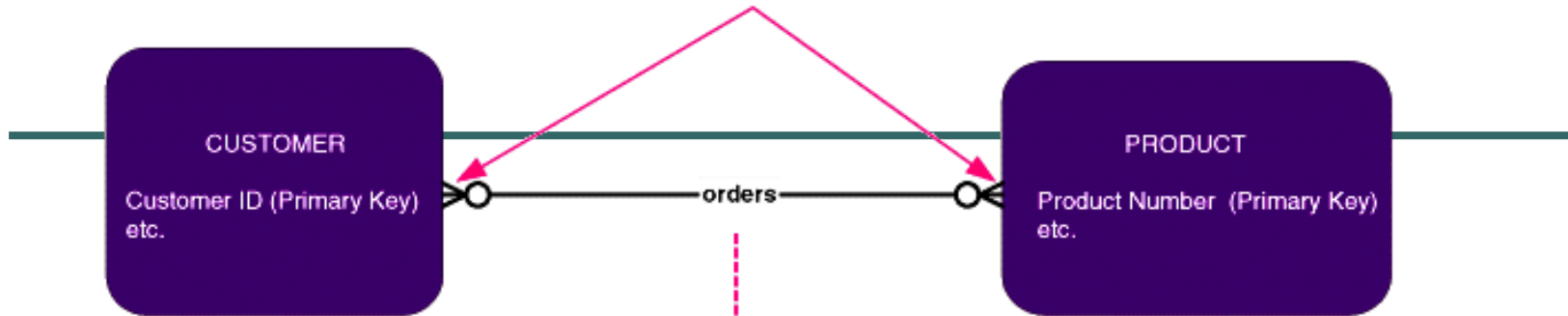
(b)



Resolving Nonspecific Relationships (continued)

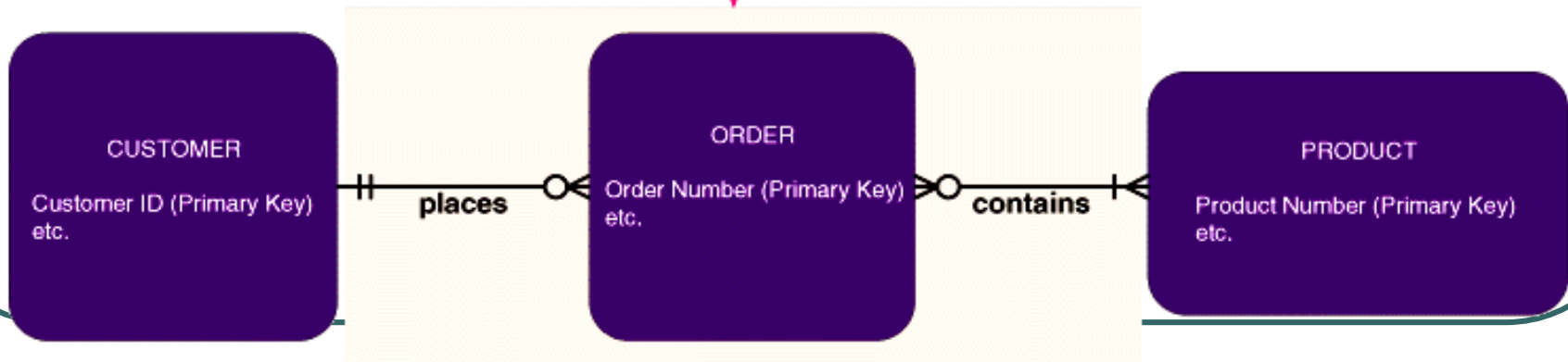
(a)

Many-to-Many Relationship



The verb or verb phrase of a many-to-many relationship sometimes suggests other entities. In this example the many-to-many is resolved by recognizing that the verb "orders" actually suggests an *event entity* called **ORDER** that relates **CUSTOMER**s to **PRODUCT**s. Notice that the new many-to-many relationship between **ORDER** and **PRODUCT** would need to be resolved.

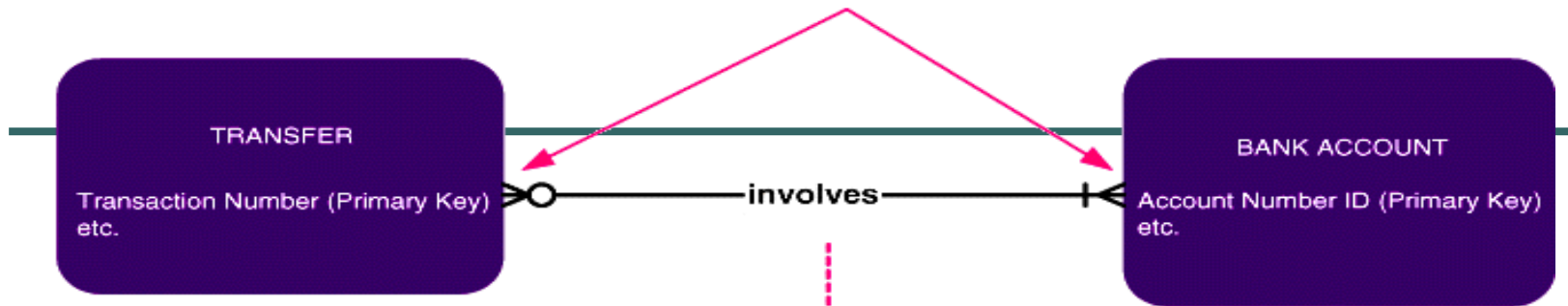
(b)



Resolving Nonspecific Relationships (continued)

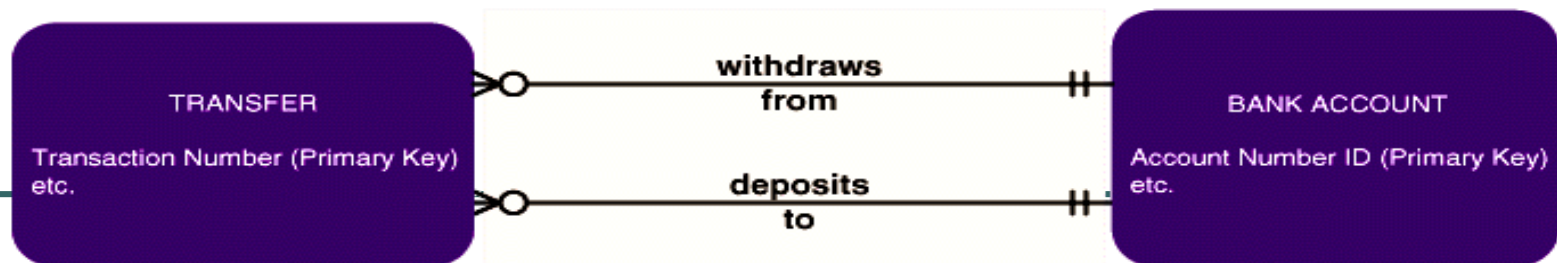
(a)

Many-to-Many Relationship



While the above relationship is a many-to-many, the many on the BANK ACCOUNT side is a known maximum of "2". This suggests that the relationship may actually represent multiple relationships ... in this case two *separate* relationships.

(b)



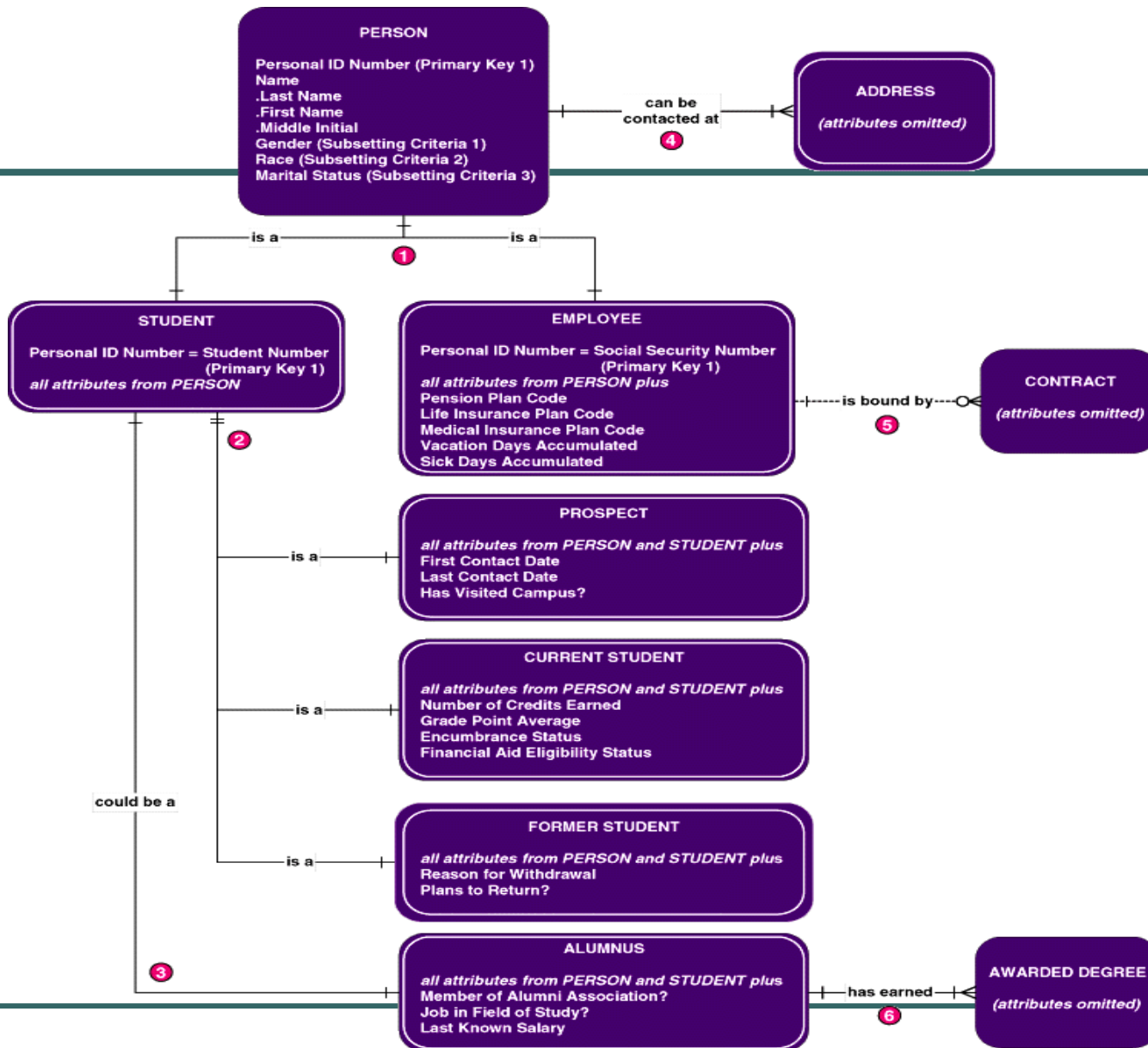
Data Modeling Concepts: Generalization

Generalization is a technique wherein the attributes that are common to several types of an entity are grouped into their own entity, called a **supertype**.

An **entity supertype** is an entity whose instances store attributes that are **common** to one or more entity subtypes.

An **entity subtype** is an entity whose instances **inherit** some common attributes from an entity supertype and then add other attributes that are unique to an instance of the subtype.

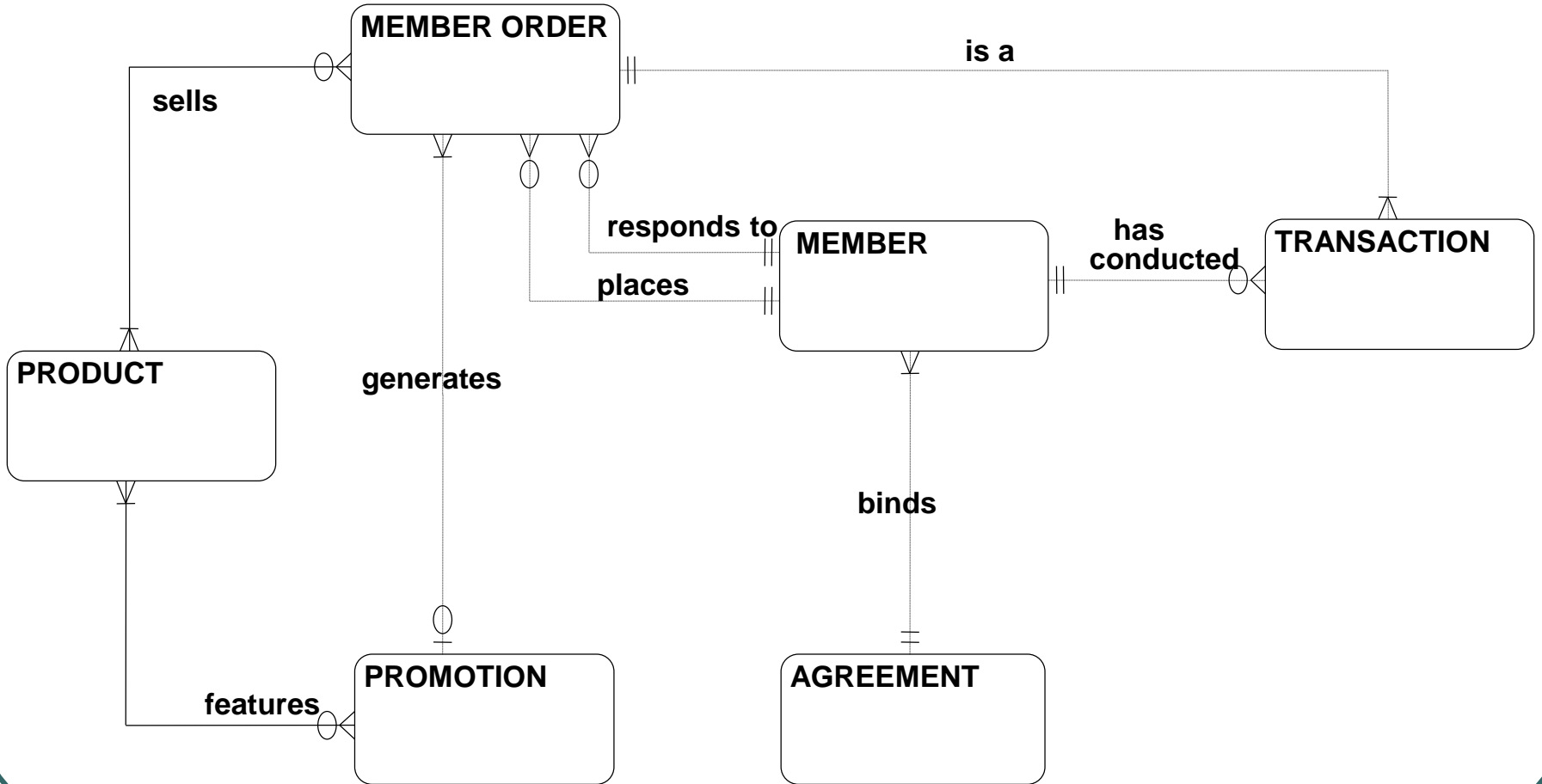
Generalization Hierarchy



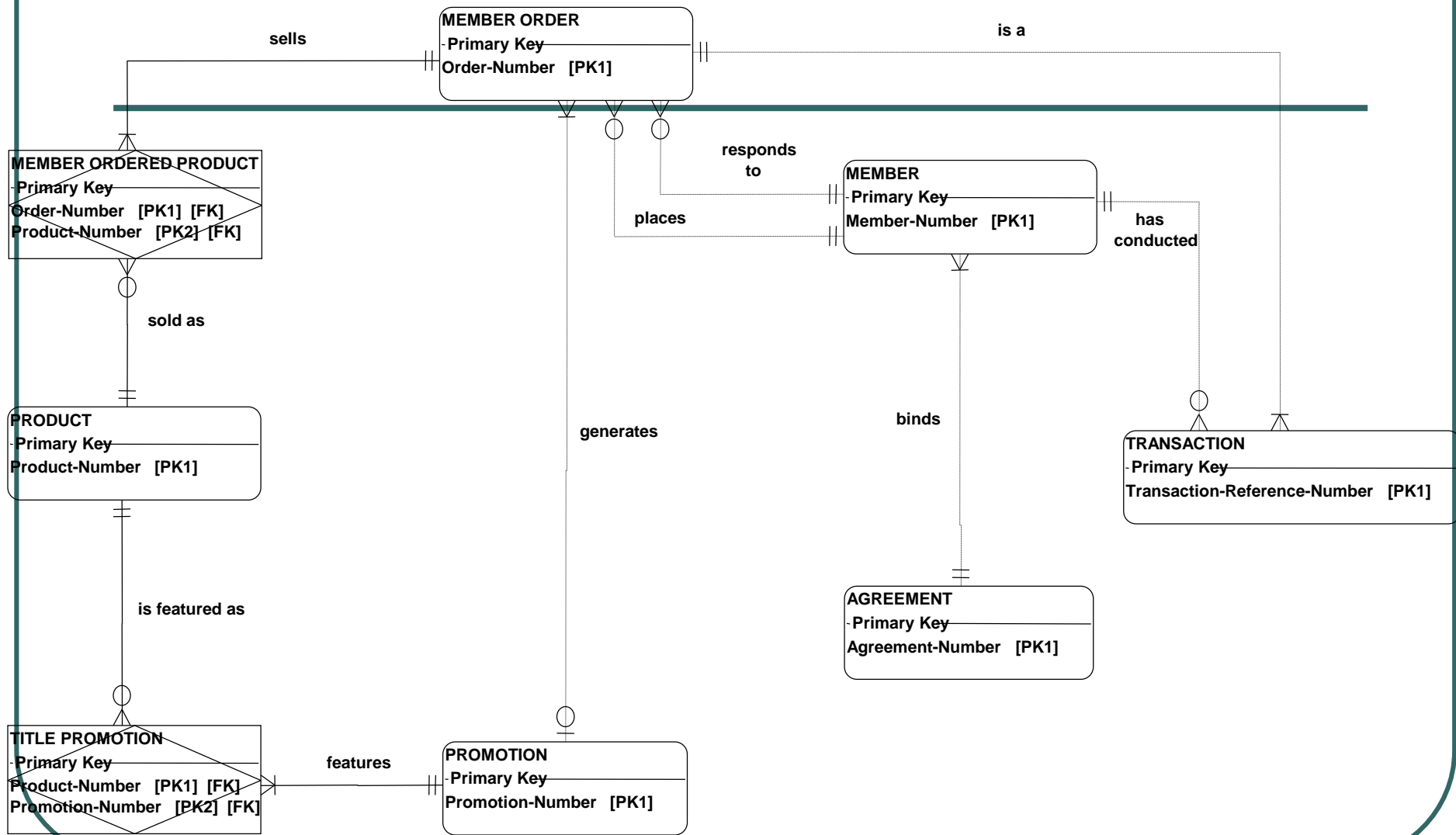
Entity Discovery for SoundStage

Entity Name	Business Definition
agreement	A contract whereby a member agrees to purchase a certain number of products within a certain time. After fulfilling that agreement, the member becomes eligible for bonus credits that are redeemable for free or discounted products.
member	An active member of one or more clubs. Note: A target system objective is to re-enroll inactive members as opposed to deleting them.
member order	An order generated for a member as part of a monthly promotion, or an order initiated by a member. Note: The current system only supports orders generated from promotions; however, customer initiated orders have been given a high priority as an added option in the proposed system.
transaction	A business event to which the Member Services System must respond.
product	An inventoried product available for promotion and sale to members. Note: System improvement objectives include (1) compatibility with new bar code system being developed for the warehouse, and (2) adaptability to a rapidly changing mix of products.
promotion	A monthly or quarterly event whereby special product offerings are made available to members.

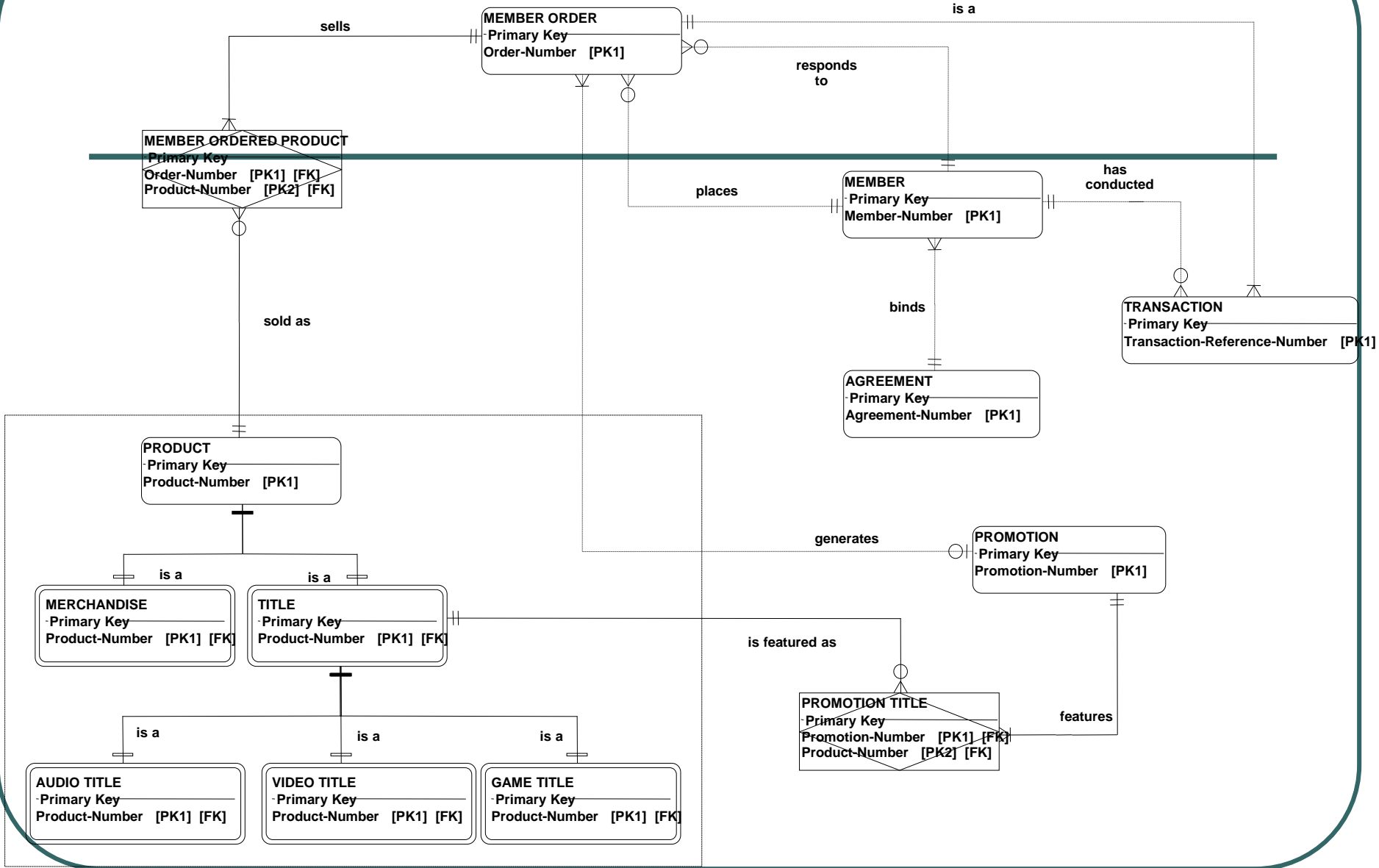
The Context Data Model



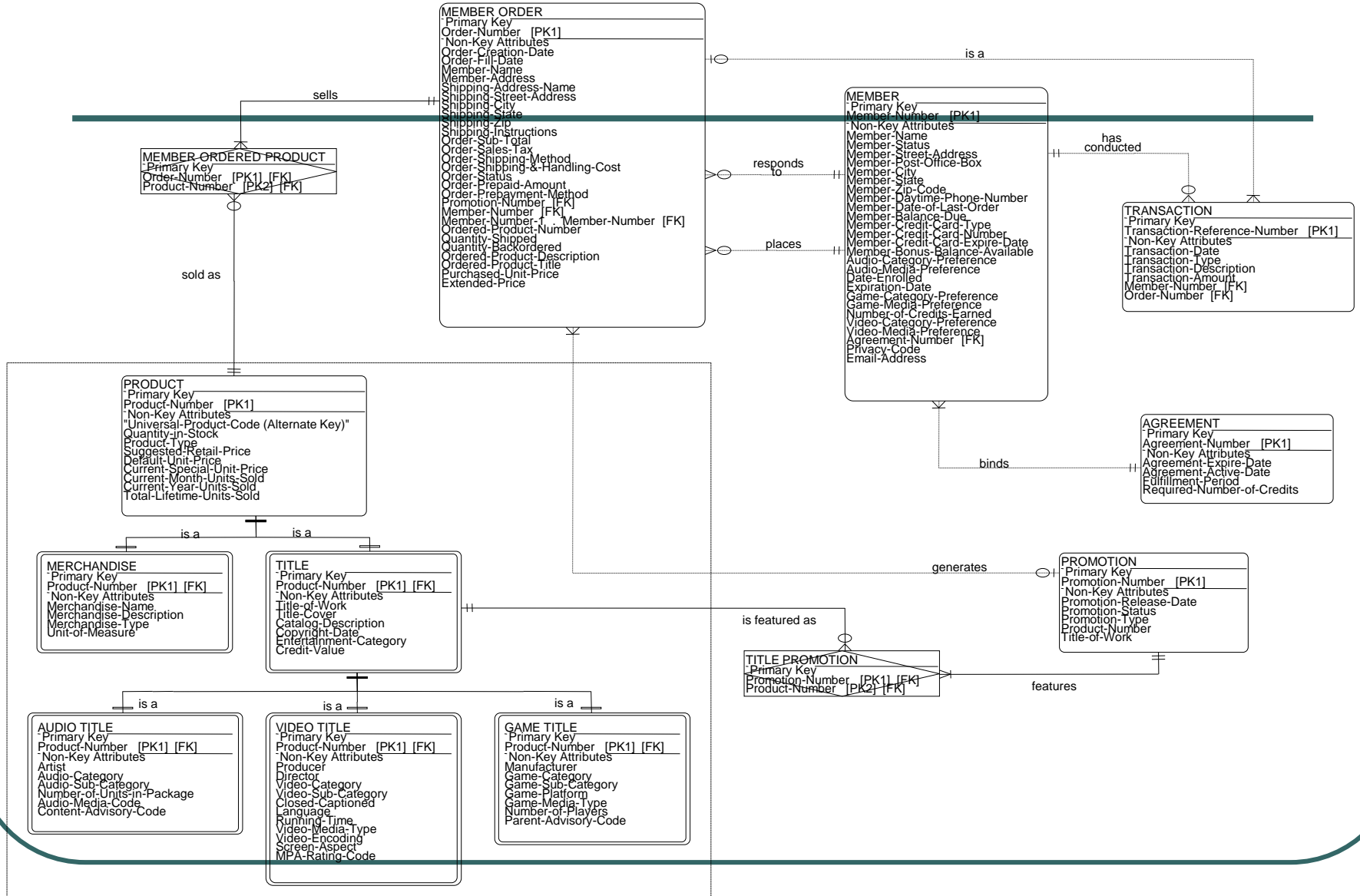
The Key-based Data Model



The Key-based Data Model With Generalization



The Fully-Attributed Data Model



Data Analysis & Normalization

Data analysis is a *process that prepares a data model* for implementation as a simple, nonredundant, flexible, and adaptable database. The specific technique is called *normalization*.

Normalization is a data analysis *technique that organizes data attributes* such that they are grouped to form *nonredundant, stable, flexible, and adaptive entities*.

Normalization: 1NF, 2NF, 3NF

- An entity is in **first normal form (1NF)** if there are **no attributes that can have more than one value** for a single instance of the entity. Any attributes that can have multiple values actually describe a separate entity, possibly an entity and relationship.
- An entity is in **second normal form (2NF)** if it is already in 1NF and if the values of **all nonprimary key attributes are dependent on the full primary key (concatenate key)**—not just part of it. Any nonkey attributes that are dependent on only part of the primary key should be moved to any entity where that partial key is actually the full key. This may require creating a new entity and relationship on the model.

Normalization: 1NF, 2NF, 3NF

- An entity is in **third normal form (3NF)** if it is already in 2NF and if the values of **its non primary key attributes are not dependent on any other non-primary key attributes**. Any non-key attributes that are dependent on other non-key attributes must be moved or deleted. Again, new entities and relationships may have to be added to the data model.

Summary

- Logical and Physical **system models**.
- **Data modeling** and its benefits.
- Basic concepts and constructs of a data model.
- **Entity-relationship** data model.
- **Repository** or Encyclopedia.
- **Normalization**
- Mapping data requirements to business operating locations.