CHAPTER

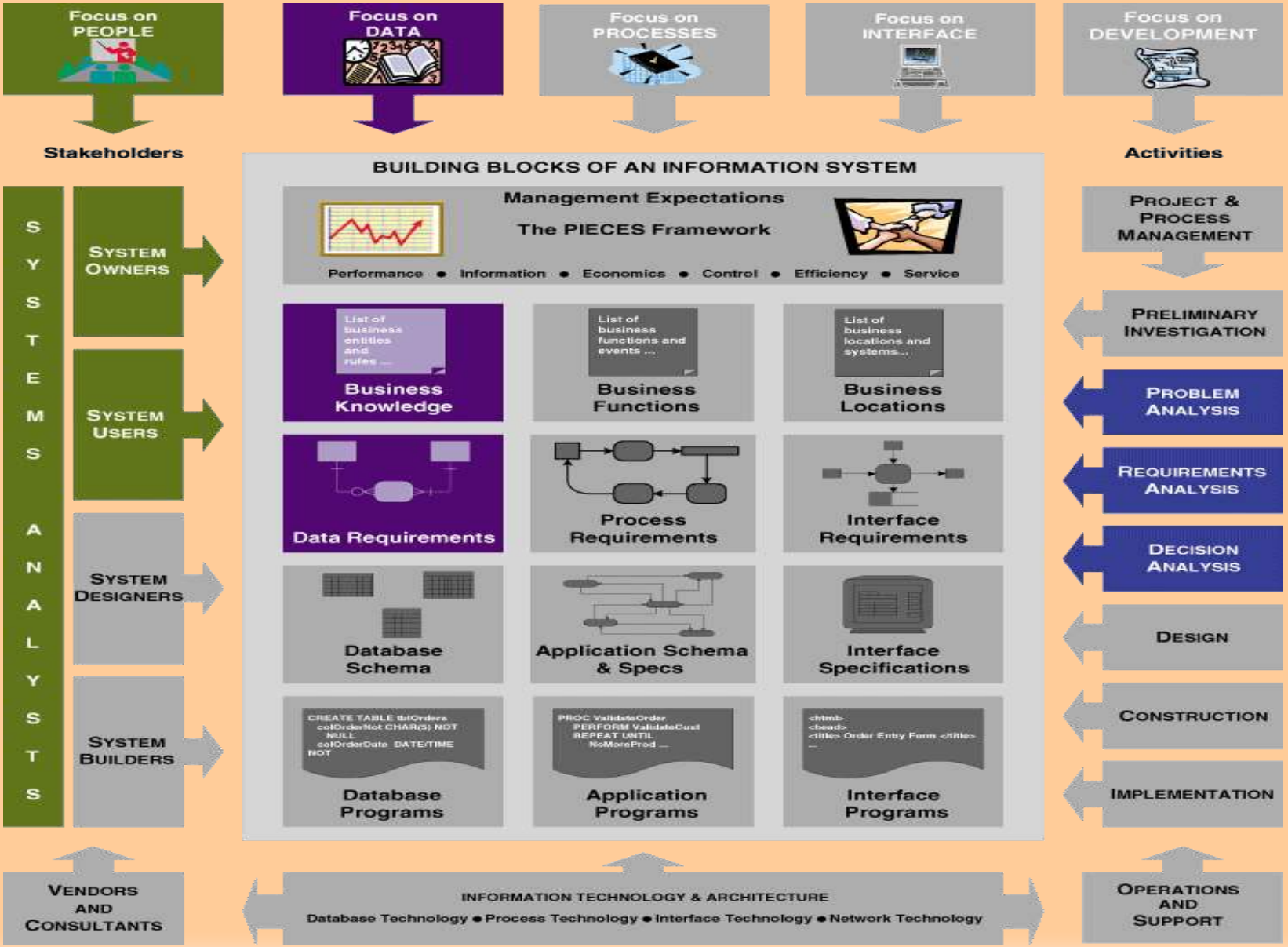
7

DATA MODELING AND ANALYSIS

Chapter Seven Data Modeling and Analysis

- Define systems modeling and differentiate between logical and physical system models.
- Define data modeling and explain its benefits.
- Recognize and understand the basic concepts and constructs of a data model.
- Read and interpret an entity relationship data model.
- Explain when data models are constructed during a project and where the models are stored.
- Discover entities and relationships.
- Construct an entity-relationship context diagram.
- Discover or invent keys for entities and construct a key-based diagram.
- Construct a fully attributed entity relationship diagram and describe all data structures and attributes to the repository or encyclopedia.
- Normalize a logical data model to remove impurities that can make a database unstable, inflexible, and nonscalable.
- Describe a useful tool for mapping data requirements to business operating locations.

Chapter Map



System Models

- A model is a representation of reality.
  - Logical models
  - Physical model

- 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

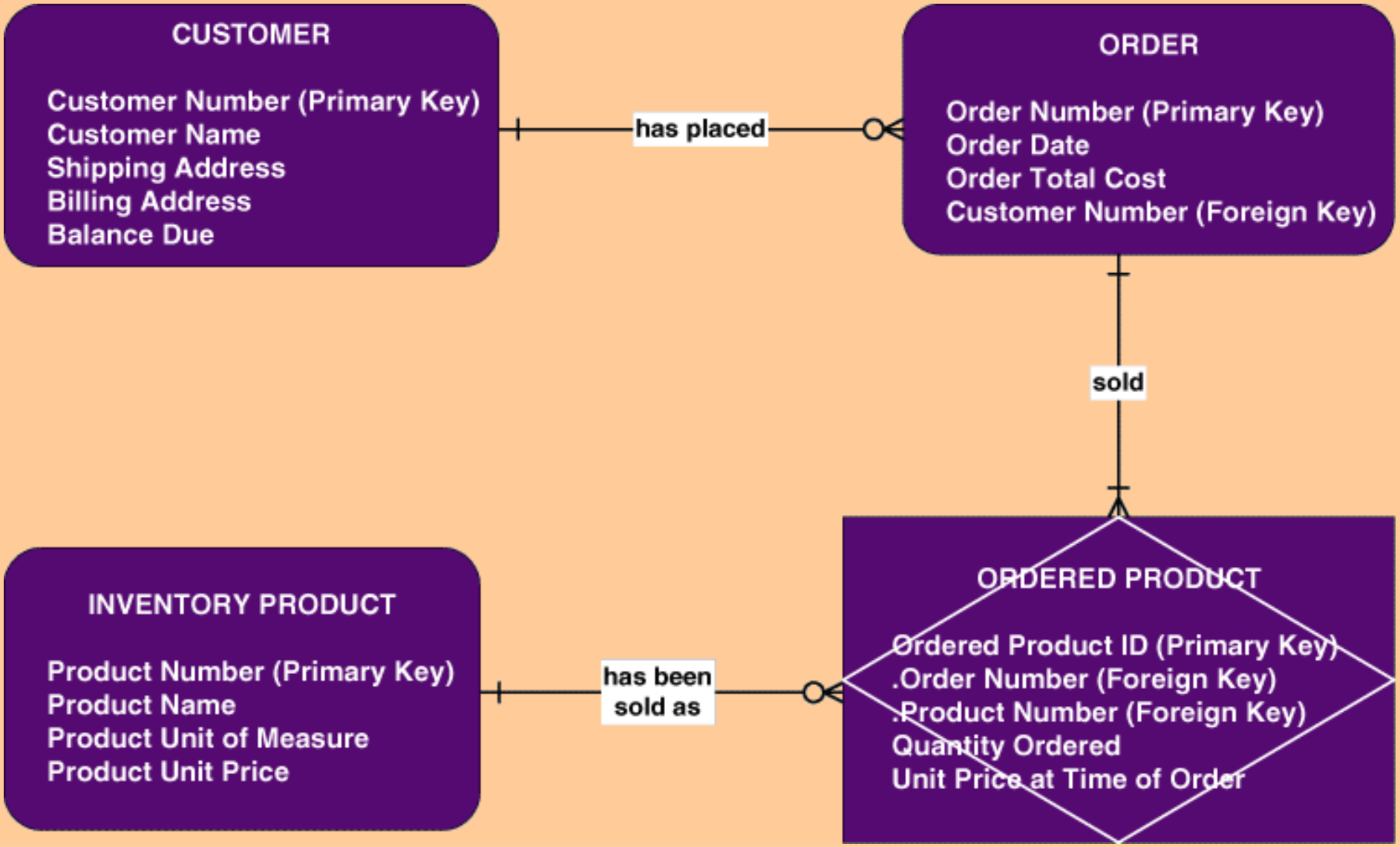
Data modeling is a technique for organizing and documenting a system's data.

Data modeling is sometimes called database modeling because a data model is eventually (finally) 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)



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.

- 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.

Data Modeling Concepts: Entity

An entity instance is a single occurrence of an entity.

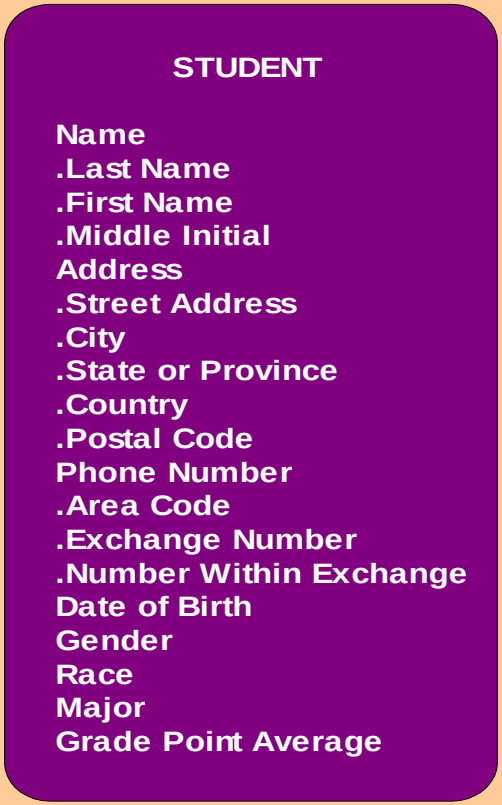
Example: instances of the entity STUDENT may include

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

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



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 (legally) 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

A key is an attribute, or a group of attributes, that assumes a unique value for each entity instance.

A group of attributes that uniquely identifies an instance of an entity is called a concatenated key.

A candidate key is a "candidate to become the primary key" of instances of an entity.

A primary key is that candidate key that will most commonly be used to uniquely identify a single entity instance.

Any candidate key that is not selected to become the primary key is called an alternate key.

A subsetting criteria is an attribute (or concatenated attribute) whose finite values divide all entity instances into useful subsets.

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

Race-
jatigosthi

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 (similarity) 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.

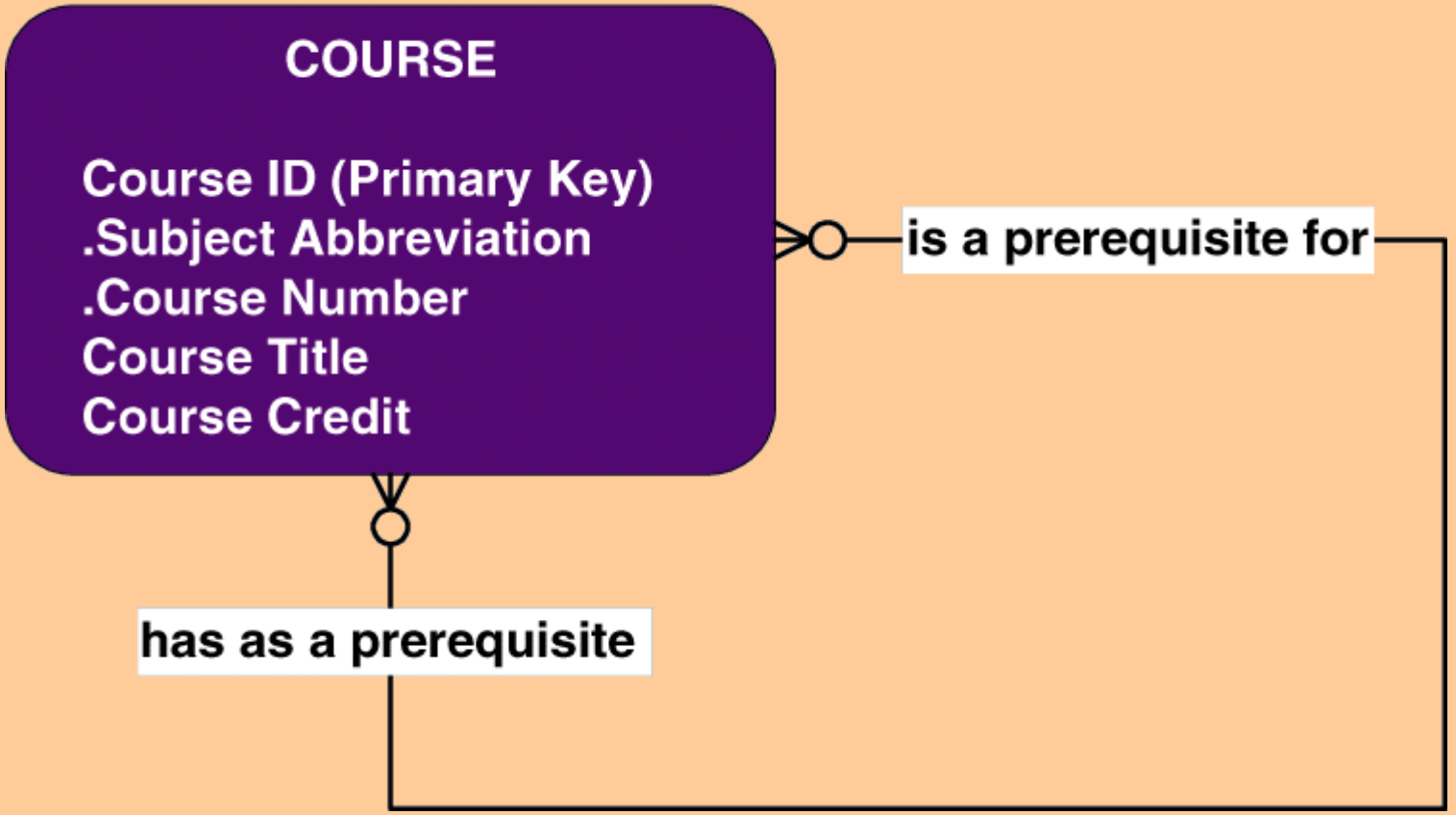
Because all relationships are bidirectional, cardinality must be defined in both directions for every relationship.

Data Modeling Concepts: Degree

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

Data Modeling Concepts: Degree

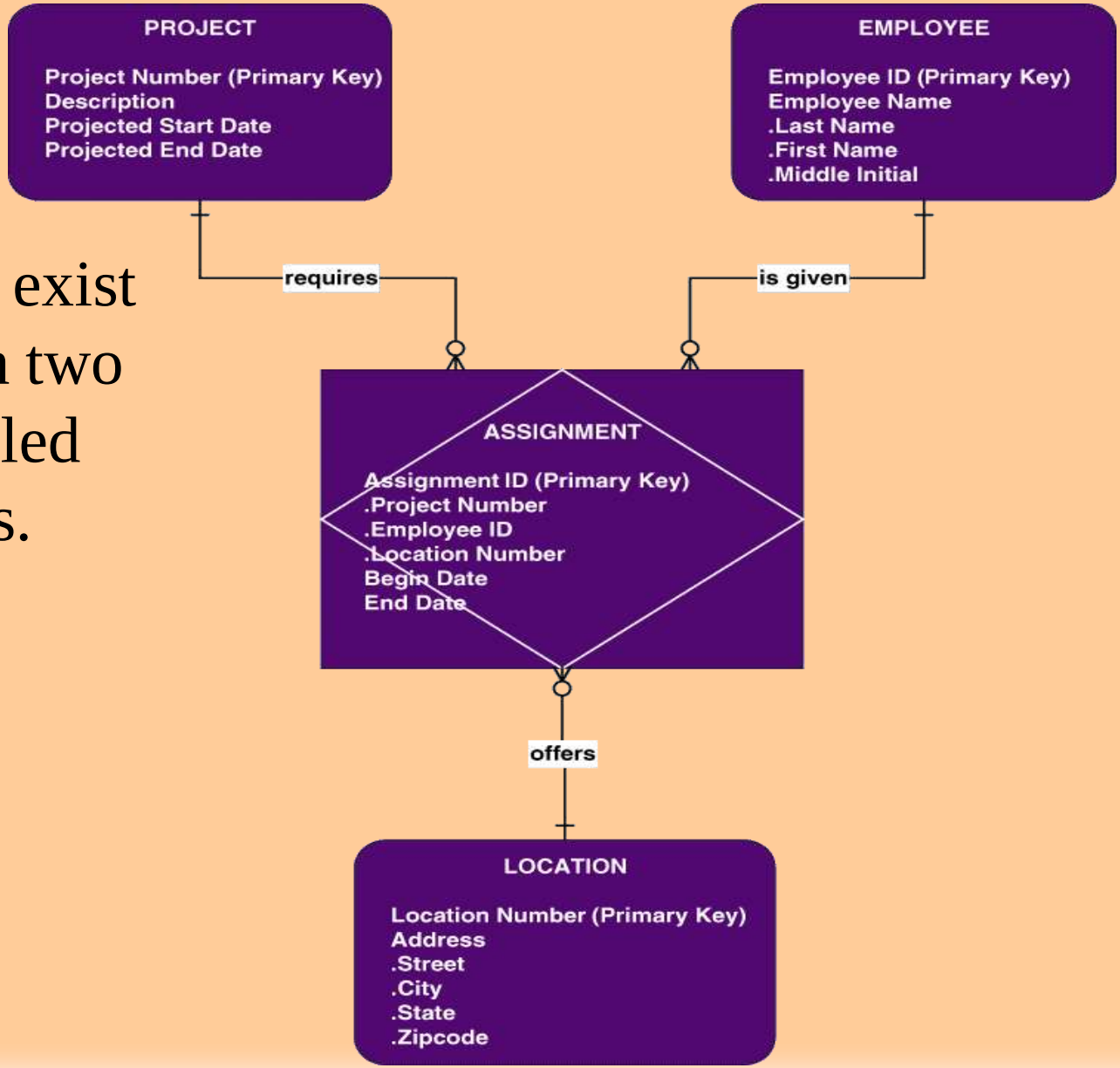
A recursive relationship is a relationship that exists between different instances of the same entity



Data Modeling Concepts: Degree

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

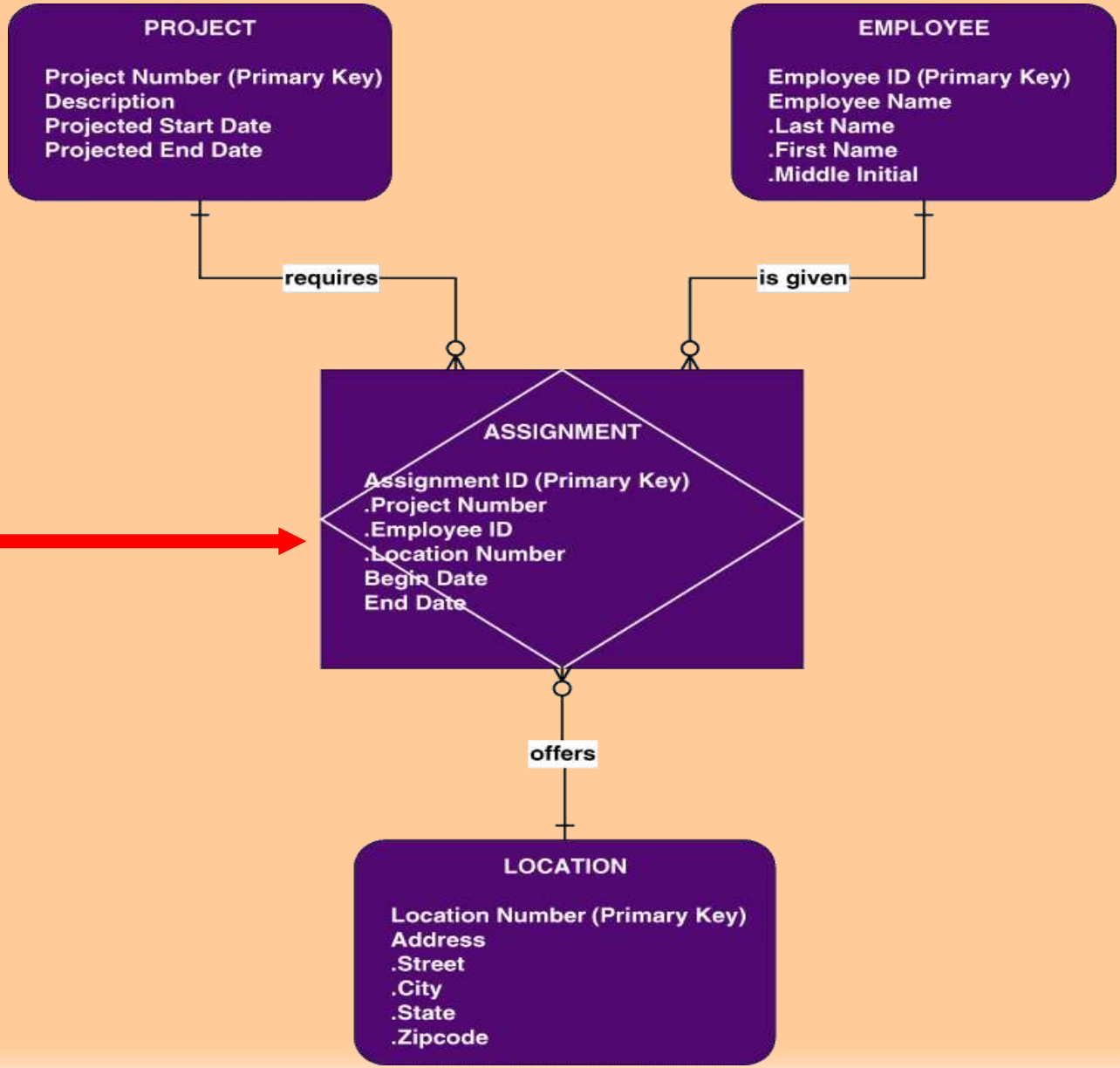
The example ERD depicts a ternary relationship.



Data Modeling Concepts: Degree

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

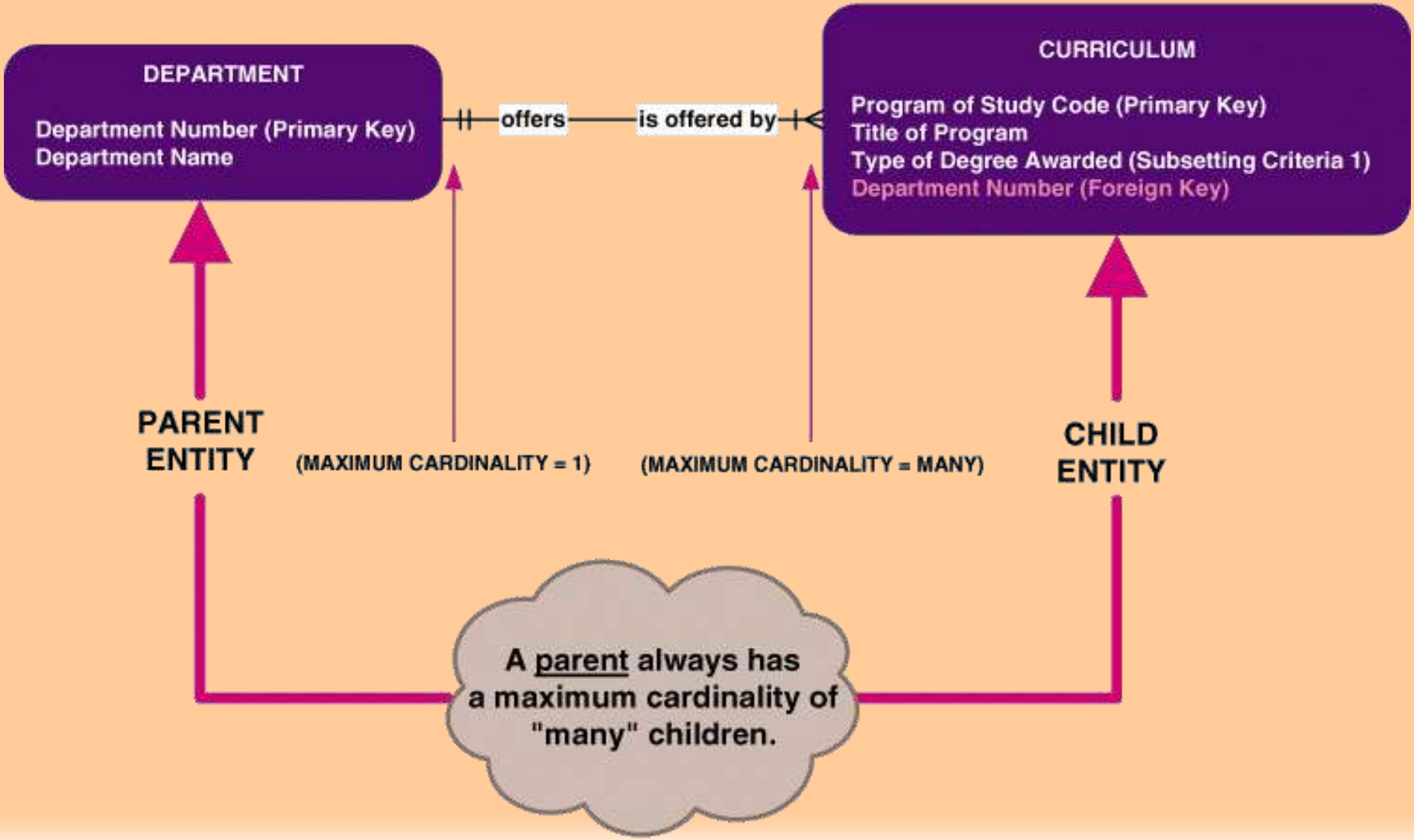
Each part of that concatenated key points to one and only one instance of each of the connecting entities.



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)

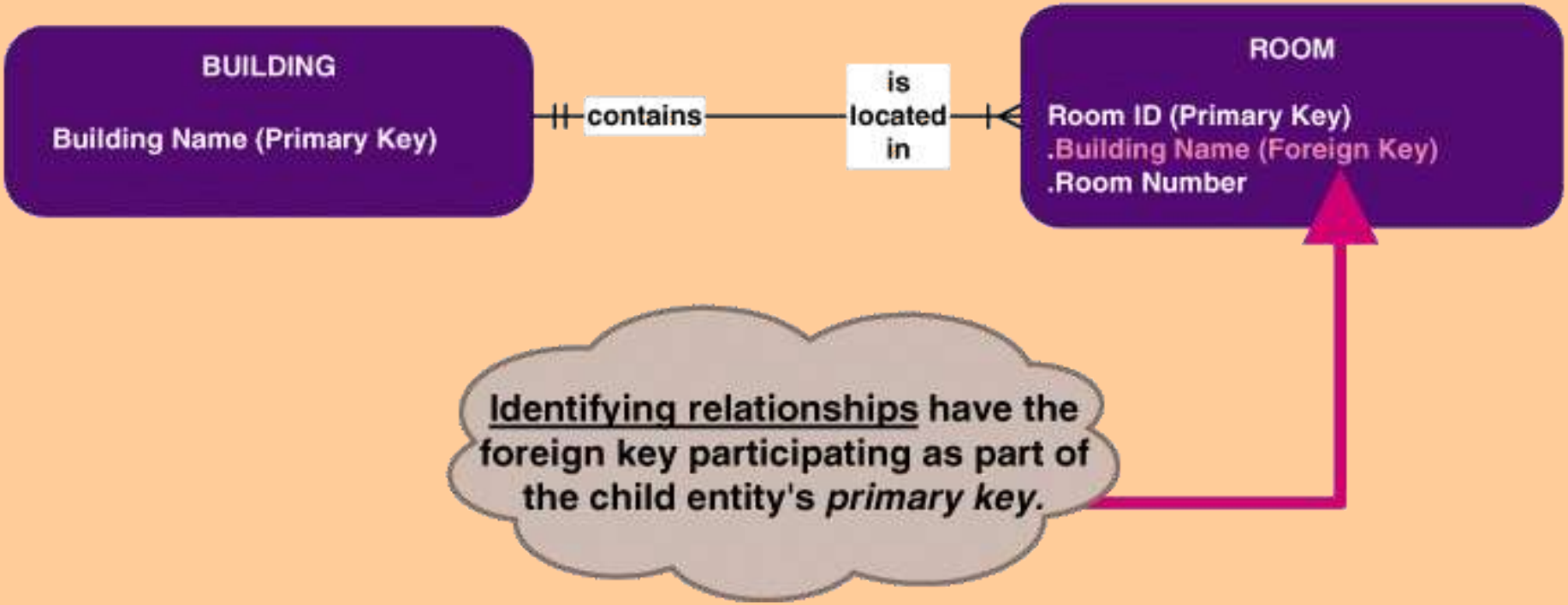


Data Modeling Concepts: Foreign Keys

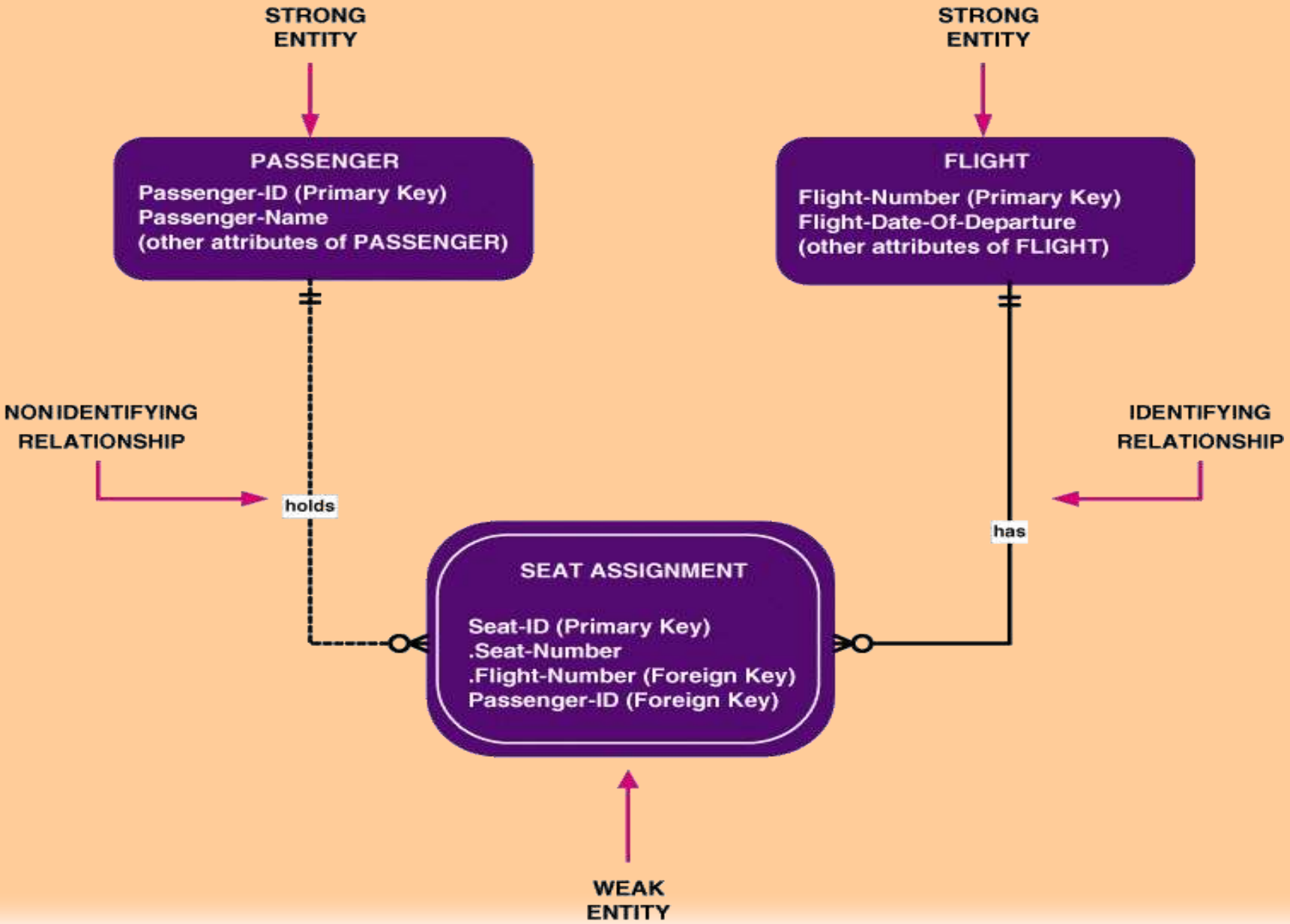
Nonidentifying relationships are those in which each of the participating entities has its own independent primary key, In other words, none of the primary key attributes is shared.

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)



Data Modeling Concepts: Foreign Keys



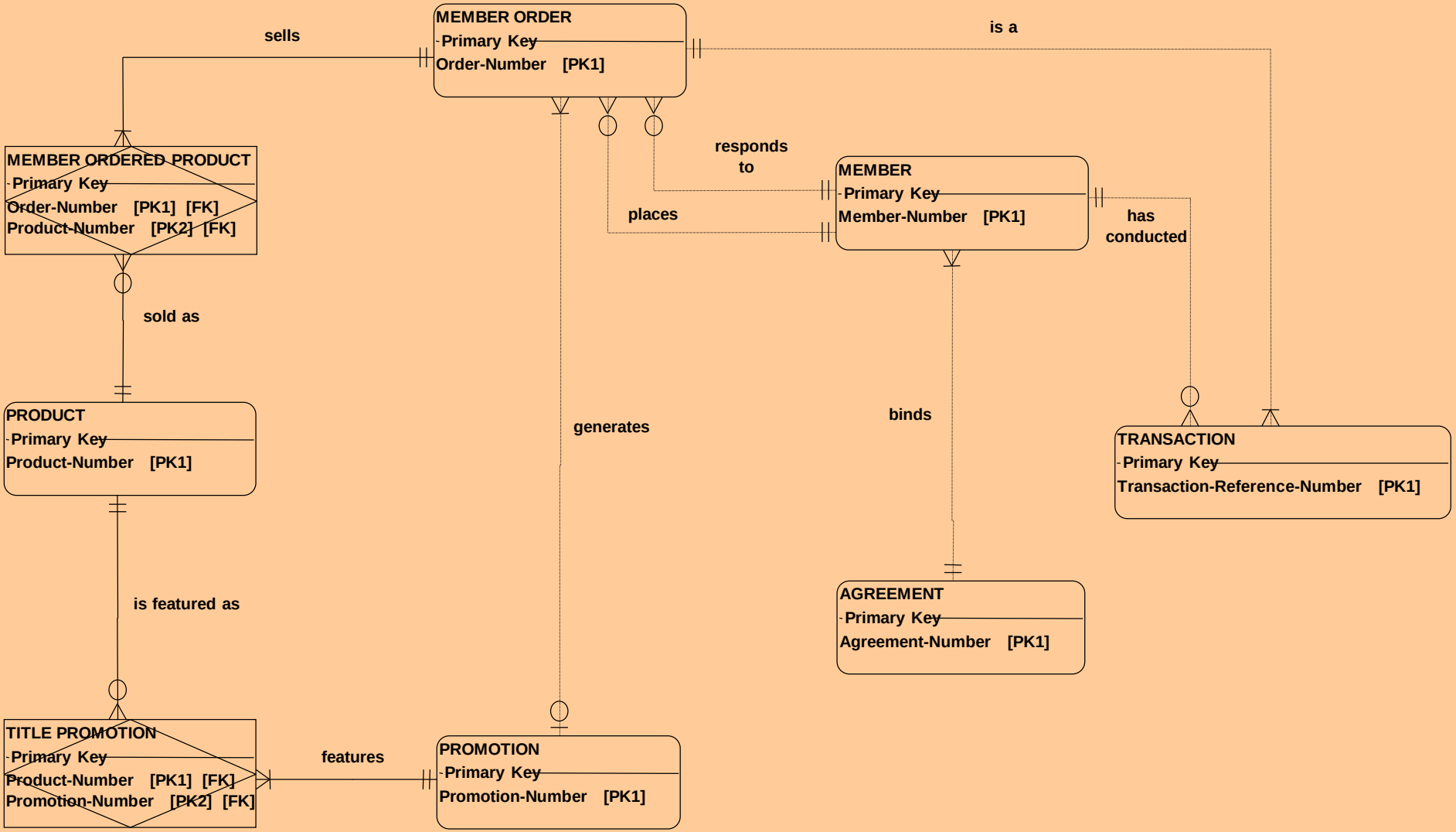
Data Modeling Concepts: Generalization

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.

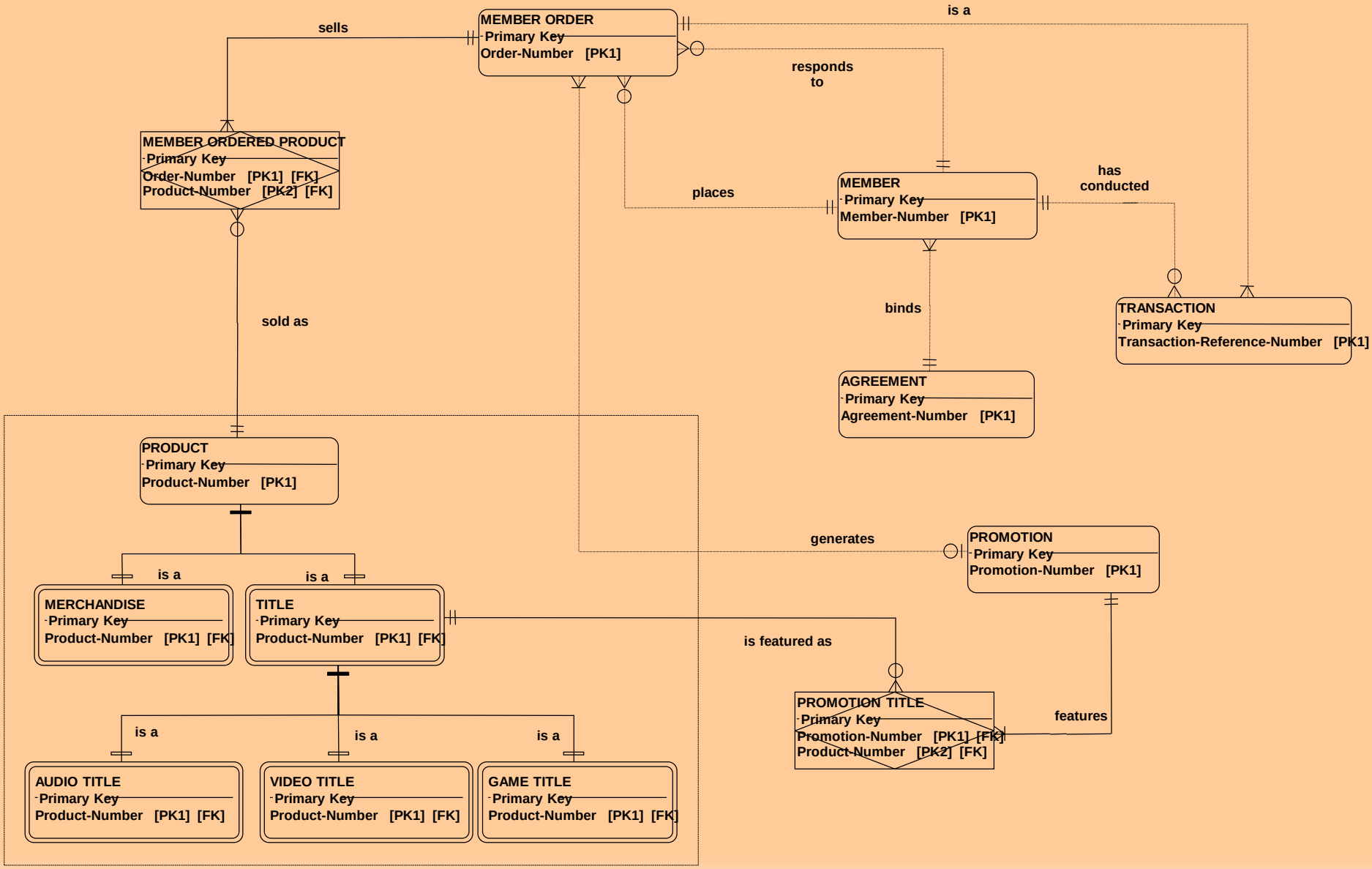
An entity supertype is an entity whose instances store attributes that are common to one or more entity subtypes.

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.

The Key-based Data Model



The Key-based Data Model With Generalization



The Fully-Attributed Data Model