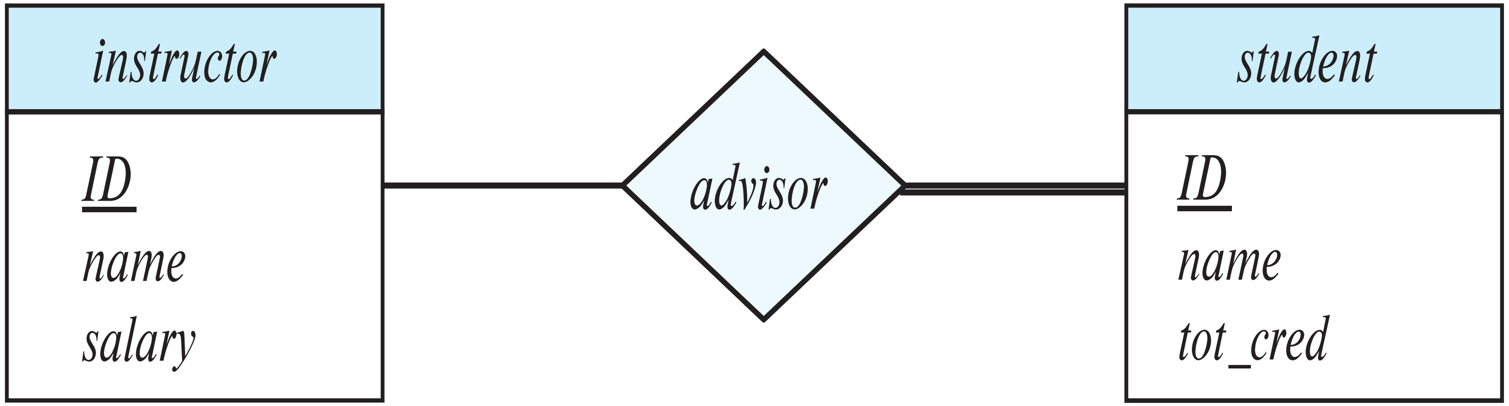
Lecture 12 - Database Modelling & Database Design (contd…)

CS211 - Introduction to Database

Participation Constraints

- **Total participation** (**indicated by double line**): every entity in the entity set participates in at least one relationship in the relationship set. It represents **NOT NULL** constraint.



  - participation of *student* in *advisor* relation is total

    Example: every *student* must have an associated instructor

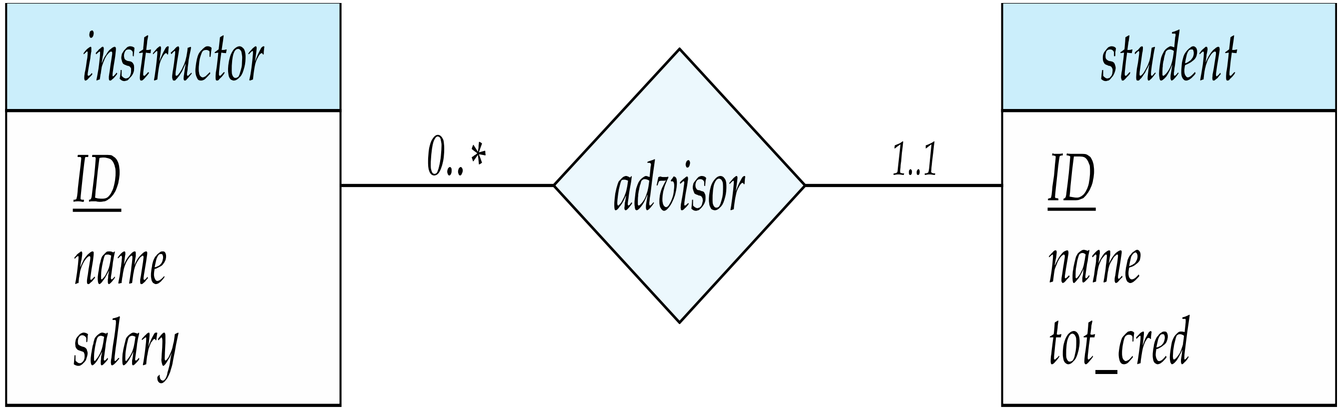
- **Partial participation**: some entities may not participate in any relationship in the relationship set. It represents the attribute values can be **NULL**.

  Example: participation of *instructor* in *advisor* is partial

Notation for Expressing More Complex Constraints

- A line may have an associated **minimum and maximum cardinality**, shown in the form ***l..h***, where ***l*** is the **minimum** and ***h*** the **maximum** cardinality
  - A minimum value of 0 indicates partial participation.
  - A minimum value of 1 indicates total participation.
  - A maximum value of 1 indicates that the entity participates in at most one relationship.
  - A maximum value of * indicates no limit.

- Example



  - Instructor can advise ZERO or more students. A student must have 1 advisor; cannot have multiple advisors

Primary Key Constraints

- Primary keys provide a way to specify how entities and relations are distinguished. We will consider:
  - Entity sets
  - Relationship sets.
  - Weak entity sets

Primary Key for Entity Sets

- By definition, individual entities are distinct.

- From database perspective, the differences among entities must be expressed in terms of their attributes.

- A **key** for an entity is a set of attributes that suffice to distinguish entities from each other.

Primary Key for Relationship Sets

- To distinguish among the various relationships of a relationship set we use the individual **primary keys** of the entities in the relationship set.
  - Let ***R*** be a relationship set involving entity sets E_1, E_2,..., E_n.
  - Let ***primary-key*(*E_i*)** denote the set of attributes that forms the primary key for entity set E_i .

Case-1: If the relationship set ***R* has no attributes** associated with it, then the set of attributes ***primary-key*(*E*$_1$)** ∪ ***primary-key*(*E*$_2$)** ∪ · · · ∪ ***primary-key*(*E*$_n$)** makes the primary key for R.

Case-2: If the relationship set ***R* has attributes** a_1, a_2,... , a_m associated with it, then the set of attributes ***primary-key*(*E*$_1$)** ∪ ***primary-key*(*E*$_2$)** ∪ ··· ∪ ***primary-key*(*E*$_n$)** ∪ **{a_1, a_2,..., a_m}** makes the primary key for R.

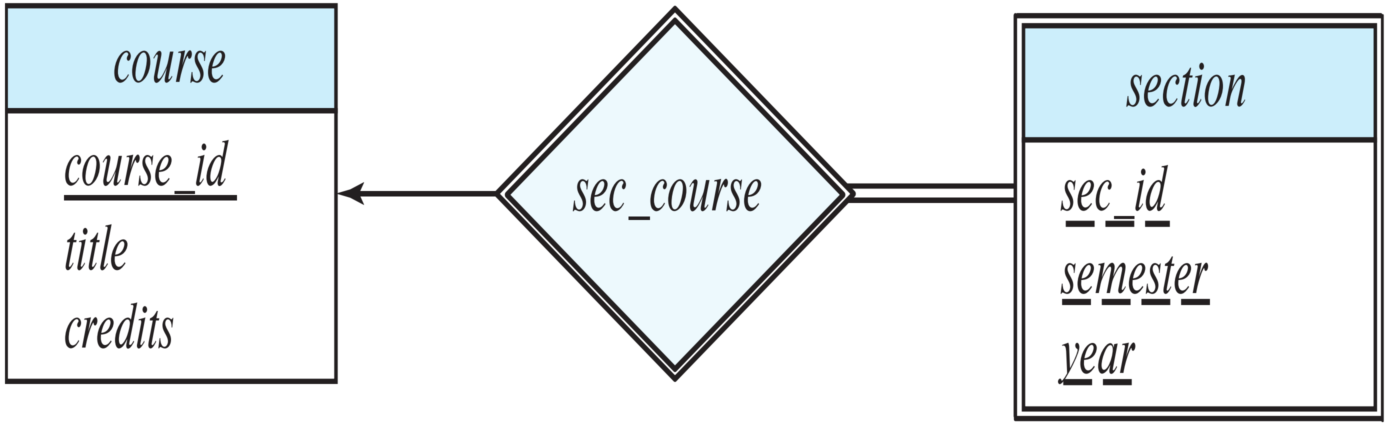
Weak Entity Sets

- An entity set that has a primary key is termed a **strong entity set**.
- An entity set that does not have sufficient attributes to form a primary key is termed a **weak entity set**.
- For a weak entity set to be meaningful, it must be associated with another entity set, called the **identifying** or **owner entity set**.
- The weak entity set is said to be **existence dependent** on the identifying entity set.
- The relationship associating the weak entity set with the identifying entity set is called the **identifying relationship**.

Weak Entity Sets (Discriminator)

- The **discriminator** of a weak entity set is a set of its attributes which combined with the **primary key** of the identifying entity set makes the primary key of weak entity set.

Primary-key of Weak entity set = Discriminator + Primary-key of owner



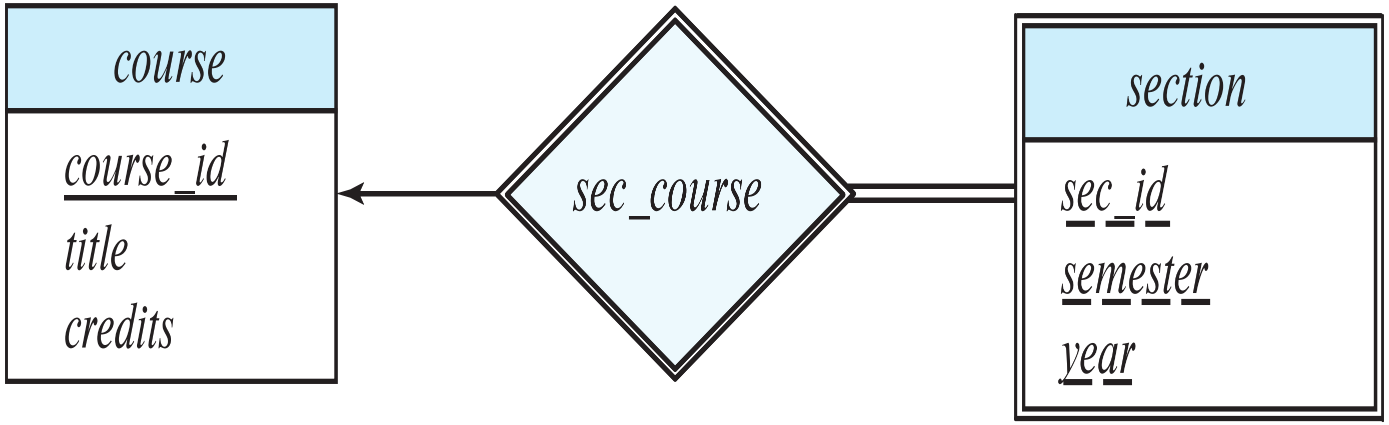
Representing Weak Entity Sets in E-R Diagram

- In E-R diagrams, a **weak entity set** is depicted via a **double rectangle**.

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

- The **identifying relationship set** connecting the weak entity set to the identifying strong entity set is depicted by a **double diamond**.

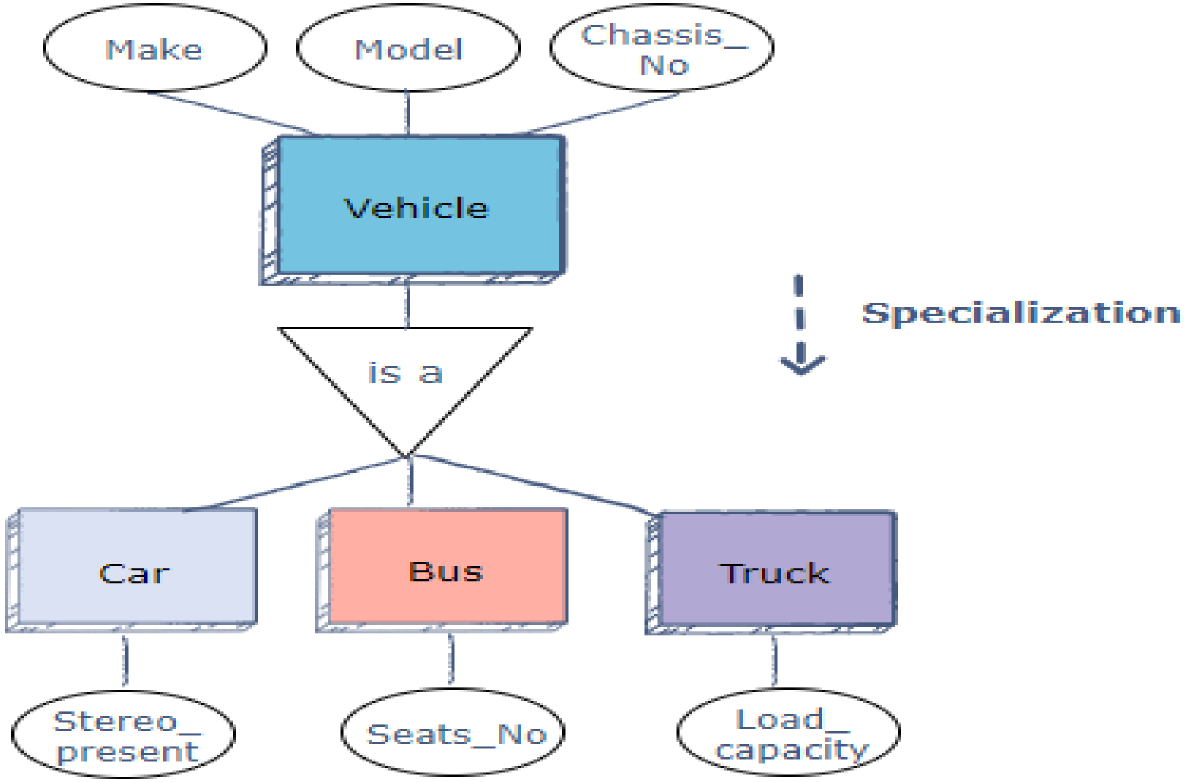
- **Primary key** for the weak entity set "*section*" is: (***course_id, sec_id, semester, year***)



Specialization

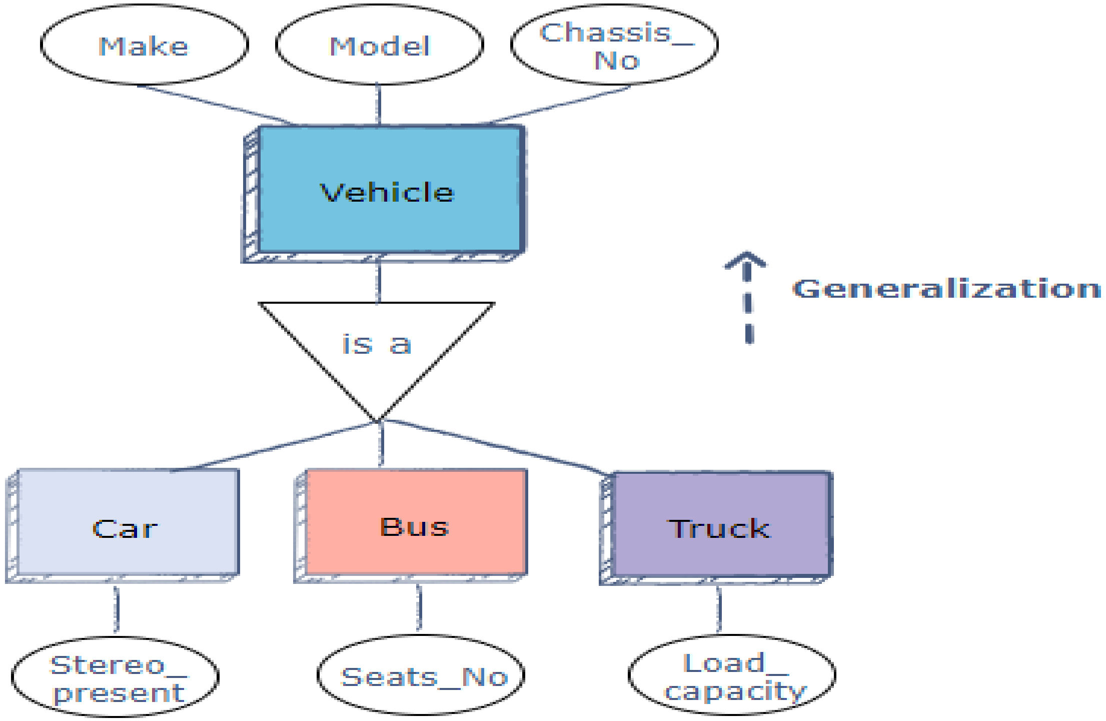
- **Specialization** is a **top-down** approach in which a higher-level entity is **divided** into multiple ***specialized*** lower-level entities.

- In addition to **sharing the attributes** of the higher-level entity, these lower-level entities have ***specific* attributes** of their own.



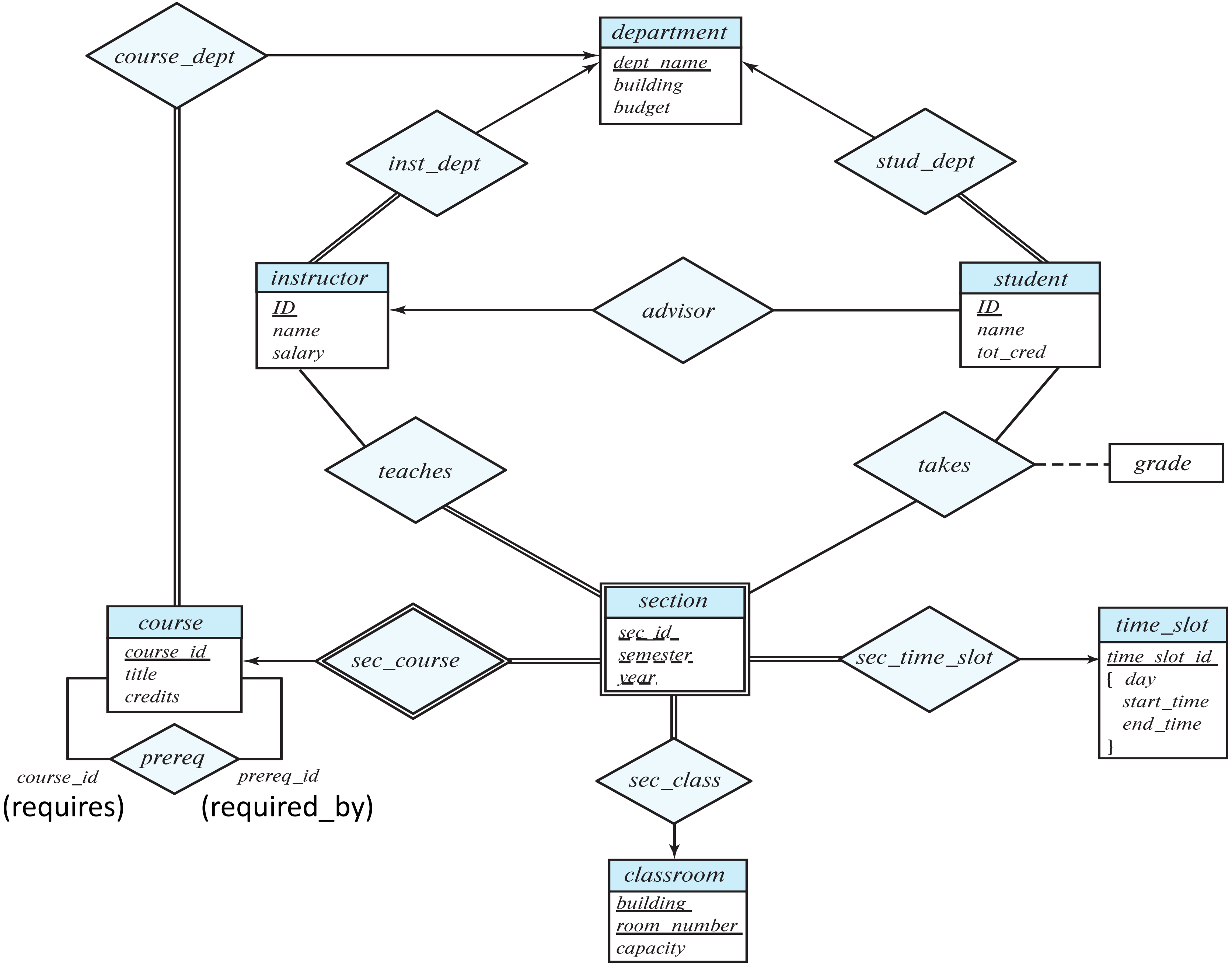
Generalization

- **Generalization** is a **bottom-up** approach in which multiple specialized lower-level entities are combined to form a single **higher-level entity**.



- Generalization is usually used to **find common attributes** among entities to form a generalized entity.

E-R Diagram for a University Enterprise



- Entity set
- Attribute type
- Entity set primary-key
- Relationship set
- Relationship set primary-key
- Relationship attribute
- Entity role
- Degree of relationship
- Cardinality constraint
- Participation constrain
- Weak entity
- Identifying relationship
- Owner entity set
- Discriminator

Database Design

Reduction to Relation Schemas

- **Entity sets** and **relationship sets** are expressed as ***relation schemas*** that represent the contents of the database.

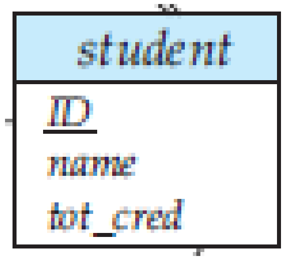
- **For each entity set and relationship set there is a unique schema that is assigned the name of the corresponding entity set or relationship set.**

- **Each schema has a number of columns (generally corresponding to attributes), which have unique names.**

Reduction of Entity Sets

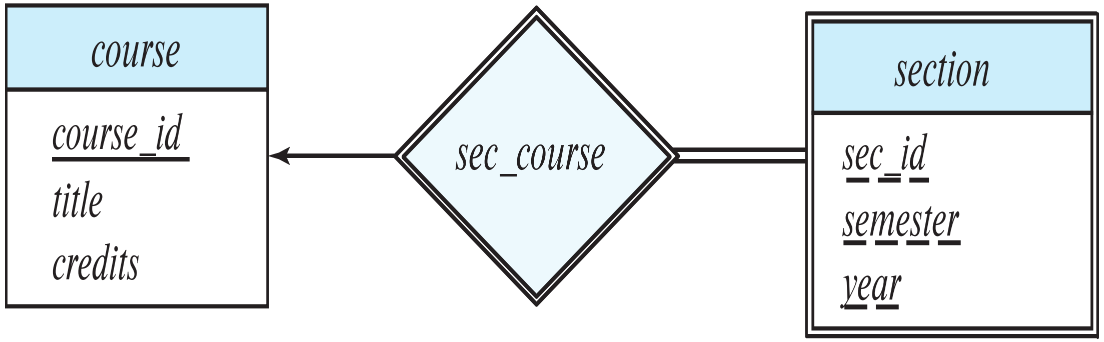
- A **strong entity set** reduces to a schema with the same attributes.

  - the **primary key** of the entity set serves as the primary key of the resulting schema



student(ID, name, tot_cred)

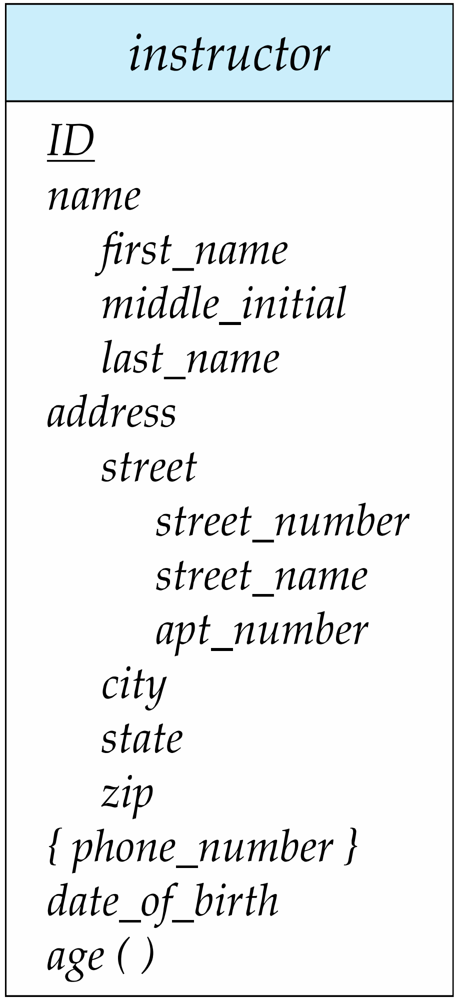
- For schemas derived from a **weak entity set**, the combination of the **primary key** of the strong entity set and the **discriminator** of the weak entity set serves as the primary key of the schema.



section (course_id, sec_id, sem, year)

Reduction of Entity Sets

- **Composite attributes** are flattened out by creating a separate attribute for each component attribute.



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 ( )*

instructor(ID,
first_name, middle_initial, last_name,
street_number, street_name, apt_number, city, state,
zip,
date_of_birth)

Reduction of Entity Sets

- A **multivalued attribute** ***M*** of an entity ***E*** is represented by a **separate schema *EM*.**

- Schema *EM* has attributes corresponding to the primary key of *E* and an attribute corresponding to multivalued attribute *M.*

- Example: Multivalued attribute ***phone_number*** of *instructor* is represented by a schema:

  ***inst_phone*= (*ID*, *phone_number*)**

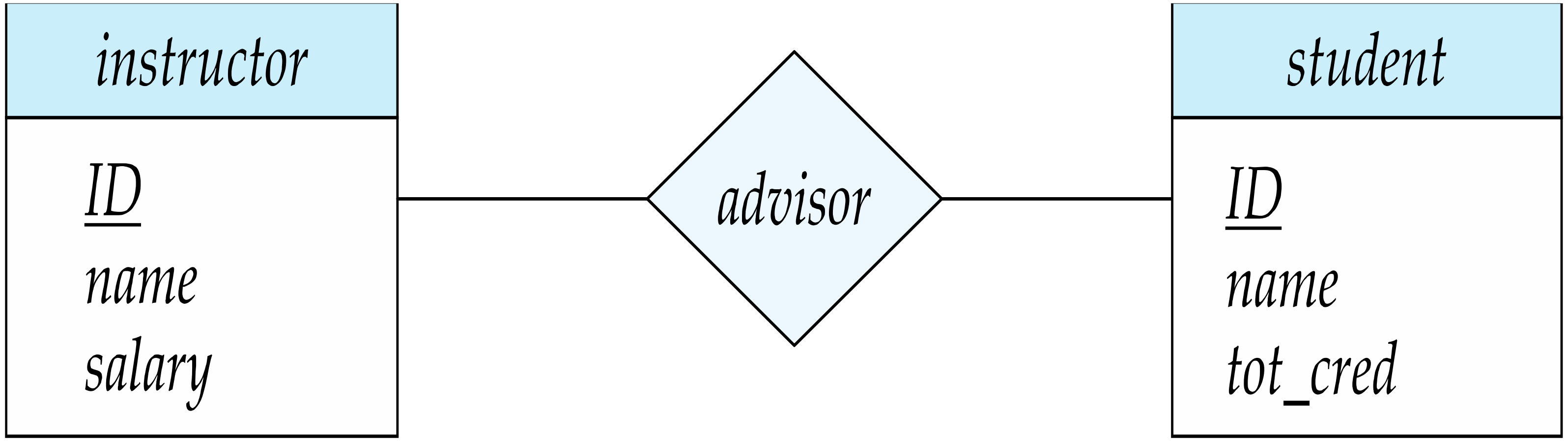
- Each value of the multivalued attribute maps to a separate tuple of the relation on schema *EM.*

  - For example, an *instructor* entity with primary key 22222 and phone numbers 456-7890 and 123-4567 maps to two tuples:

    (22222, 456-7890) and **(22222, 123-4567)**

Reduction of Relationship Sets

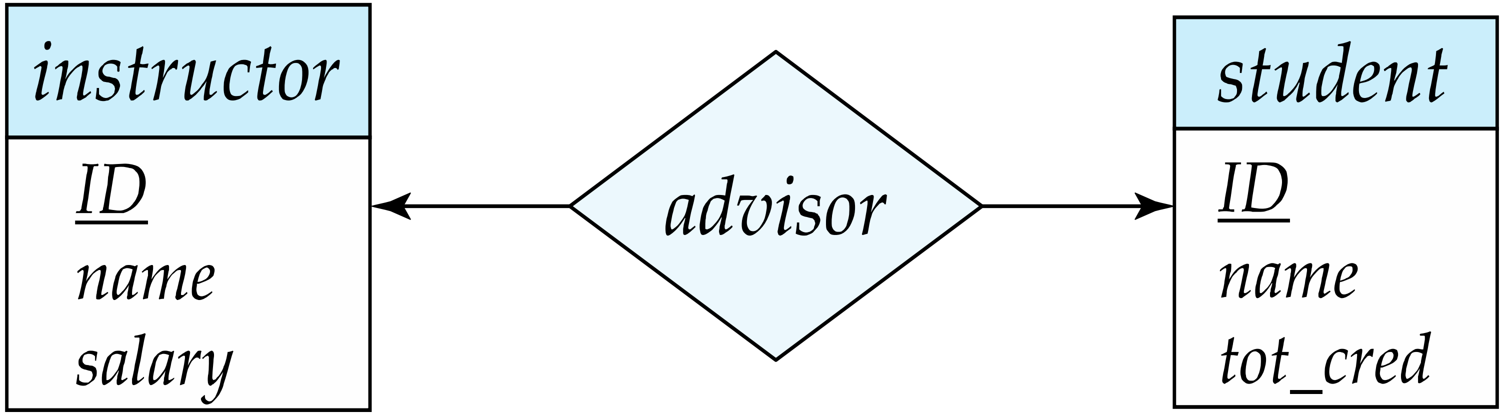
- For a **binary many-to-many relationship**, the union of the primary-key attributes of **all the participating entity sets** becomes the primary key.



***advisor* = (*i_ID, s_ID*)**

Reduction of Relationship Sets

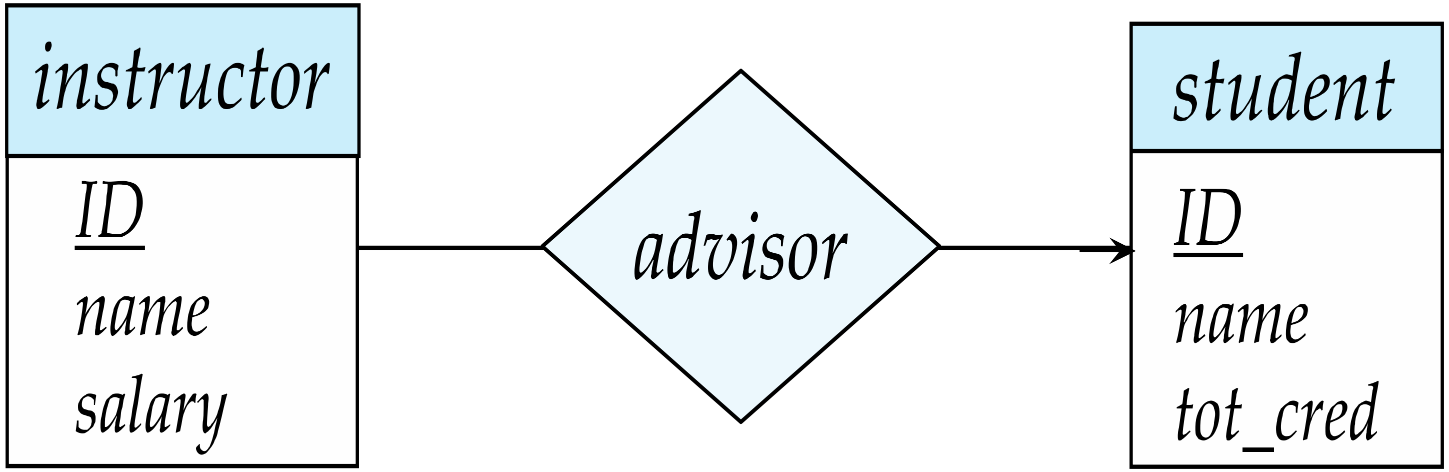
- For a **binary one-to-one relationship set**, the primary key of **either entity set** can be chosen as the primary key. The choice can be made arbitrarily.



***advisor* = (*s_ID*)**

Reduction of Relationship Sets

- For **a binary many-to-one or one-to-many relationship set**, the primary key of the **entity set on the "many" side** of the relationship set serves as the primary key.



***advisor* = (*i_ID*)**

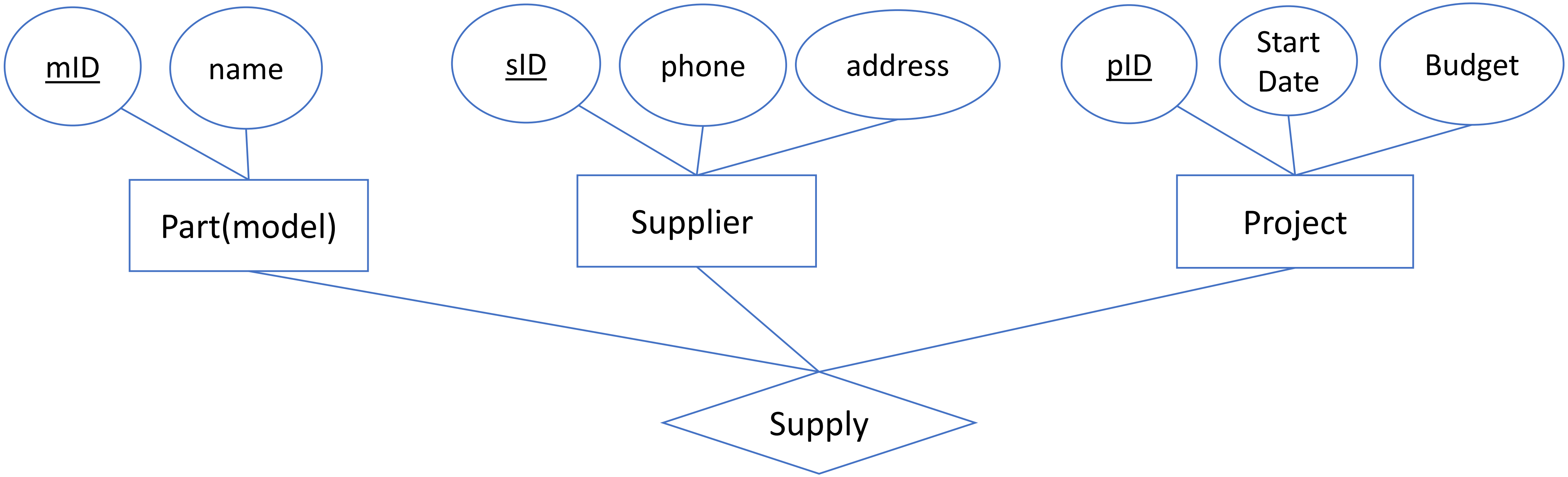
Reduction of Relationship Sets

- For **an *n*-ary relationship set without any arrows on its edges**, the union of the primary key-attributes of **all the participating entity sets** becomes the primary key.

- For an ***n*-ary relationship set with an arrow on one of its edges**, the primary keys of the **entity sets not on the "arrow" side** of the relationship set serve as the primary key for the schema.

Reduction of Relationship Sets

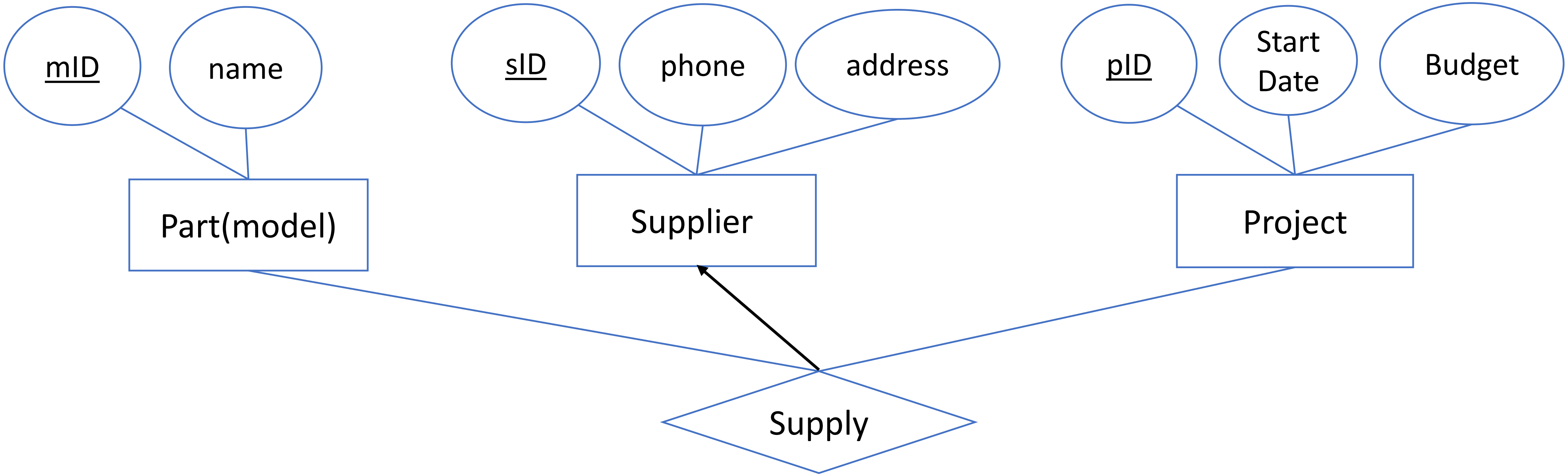
Multi-degree relationship → schema (Case-1)



supply(<u>mID</u>, <u>sID</u>, <u>pID</u>)

Reduction of Relationship Sets

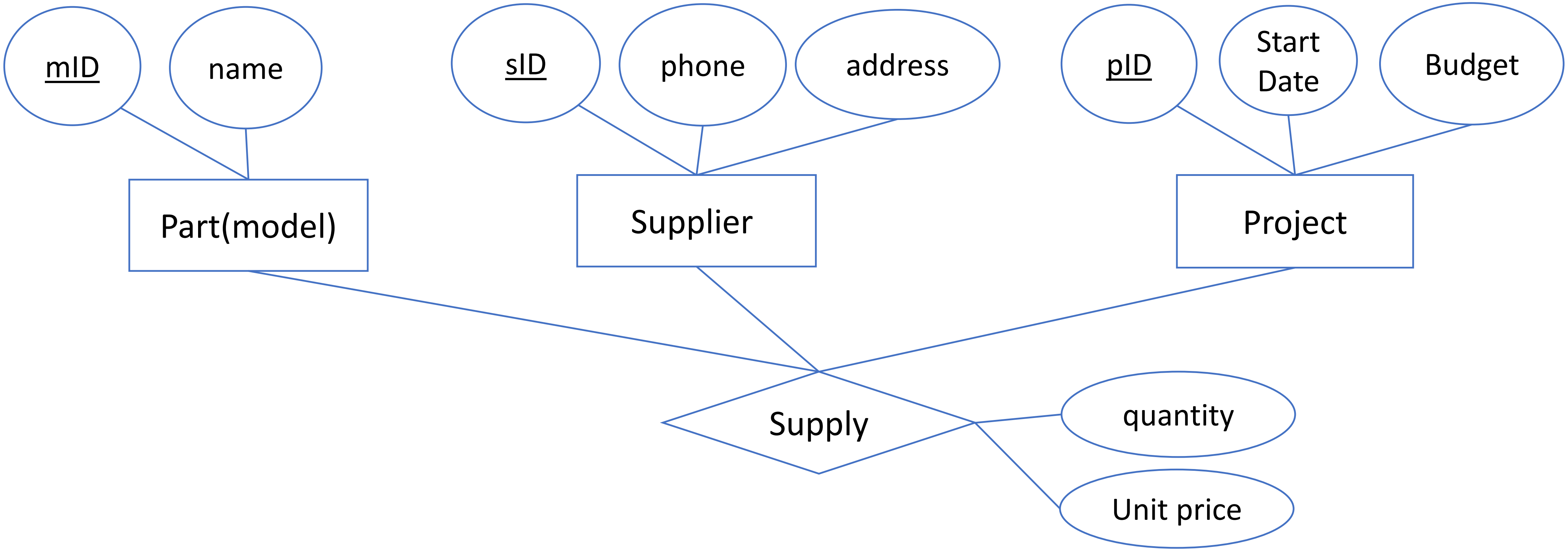
Multi-degree relationship → schema (Case-2)



supply(mID, pID)

Reduction of Relationship Sets

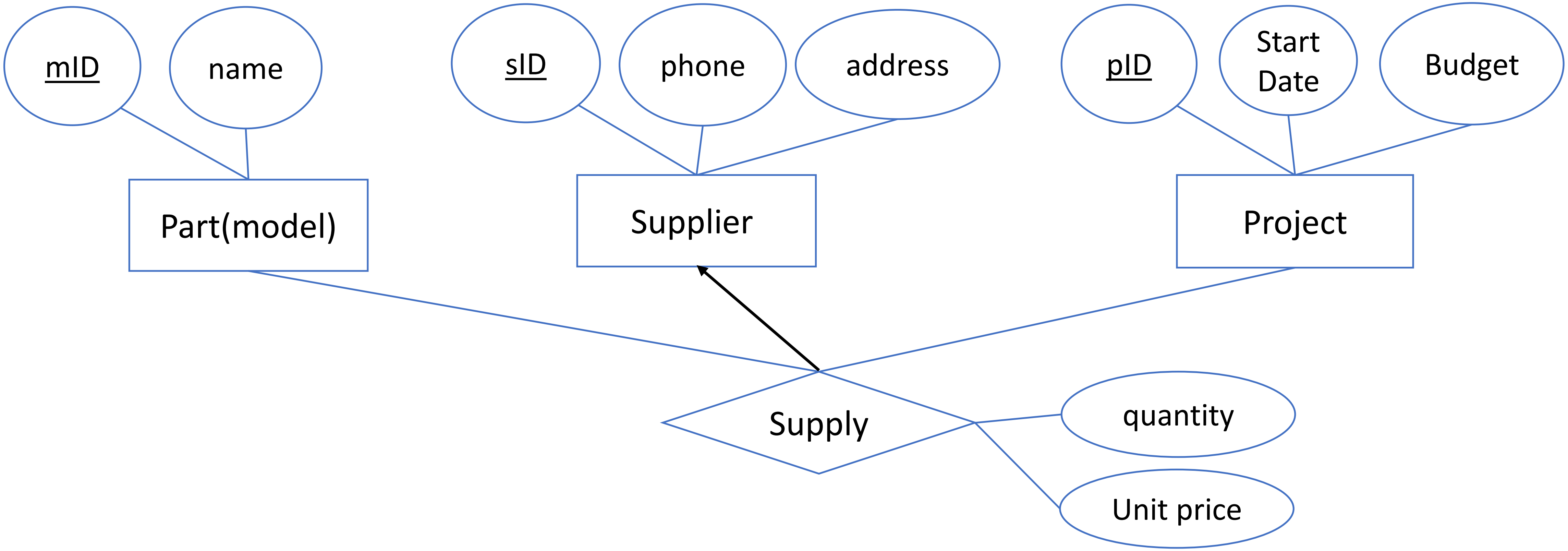
Multi-degree relationship → schema (Case-3)



supply(<u>mID</u>, <u>sID</u>, <u>pID</u>, <u>quantity</u>, <u>unitprice</u>)

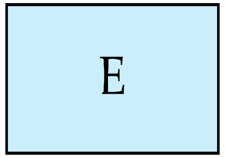
Reduction of Relationship Sets

Multi-degree relationship → schema (Case-4)

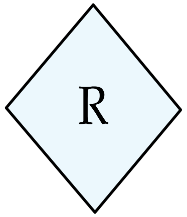


supply(mID, pID, quantity, unitprice)

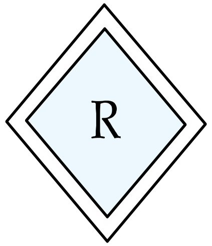
Summary of Symbols Used in E-R Notation



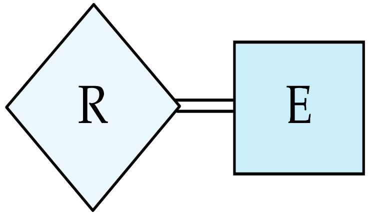
entity set



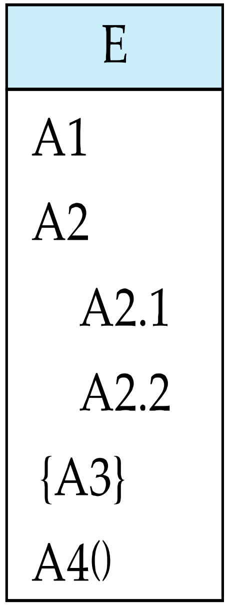
relationship set



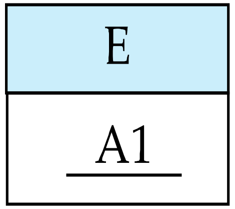
identifying
relationship set
for weak entity set



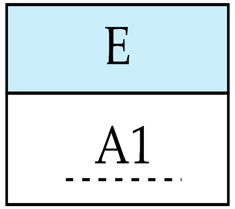
total participation
of entity set in
relationship



attributes:
simple (A1),
composite (A2) and
multivalued (A3)
derived (A4)

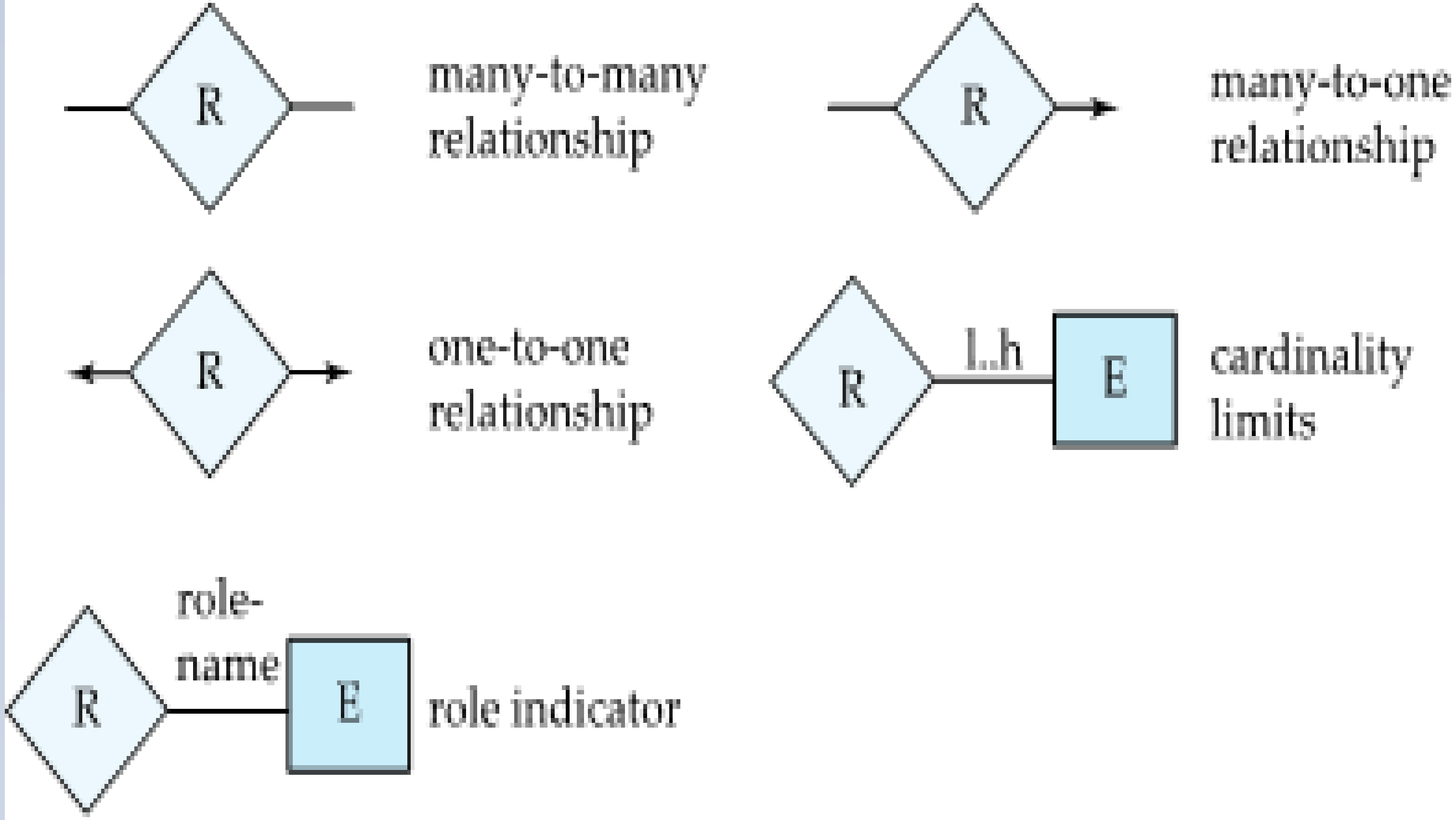


primary key



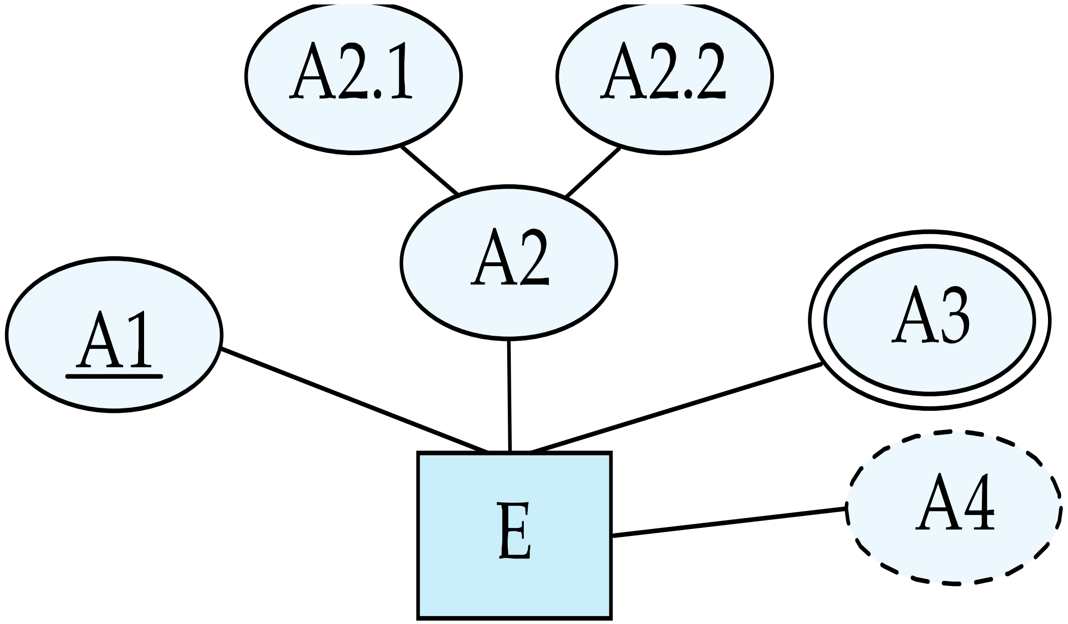
discriminating
attribute of
weak entity set

Summary of Symbols Used in E-R Notation

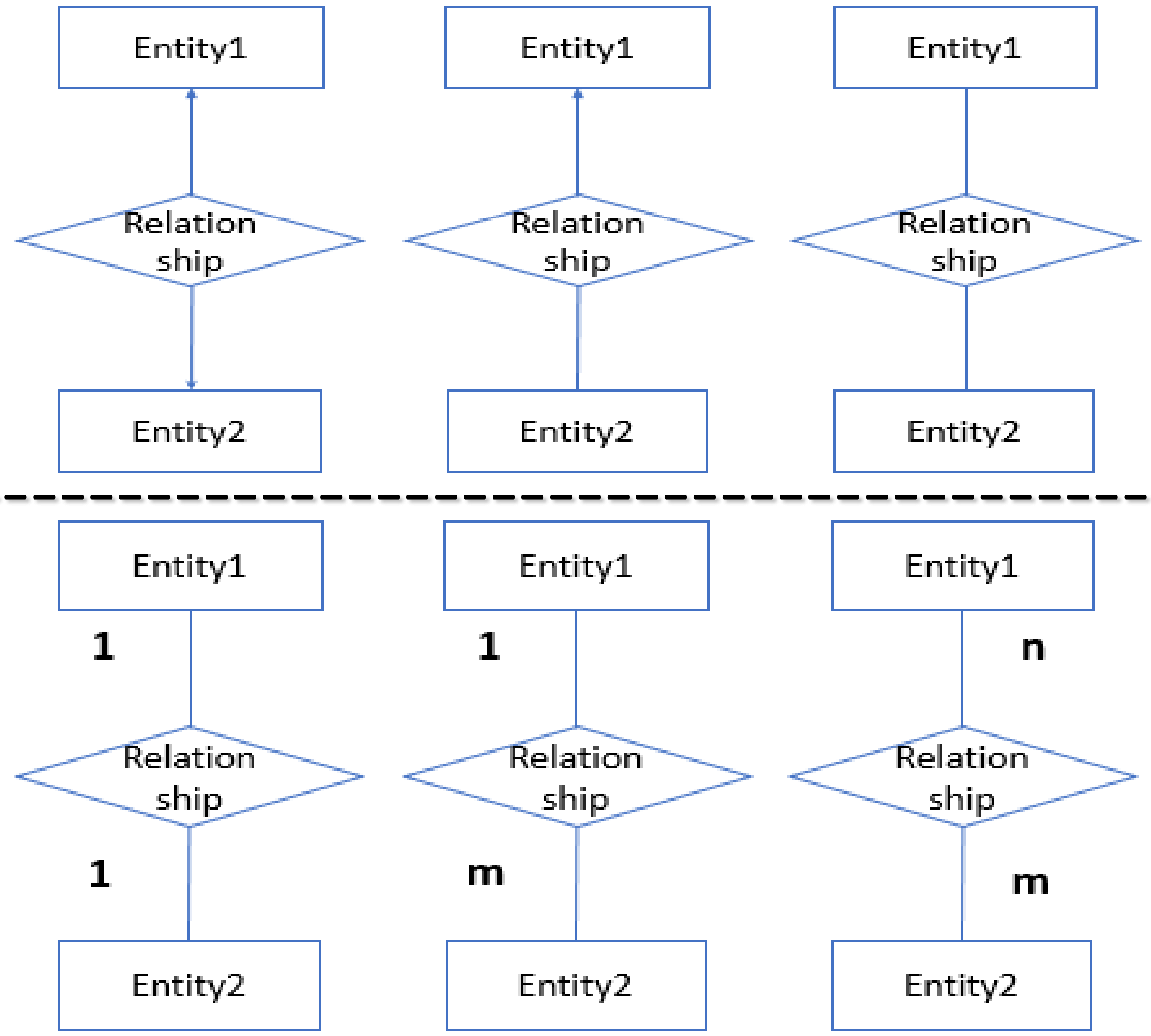


Alternative ER Notations

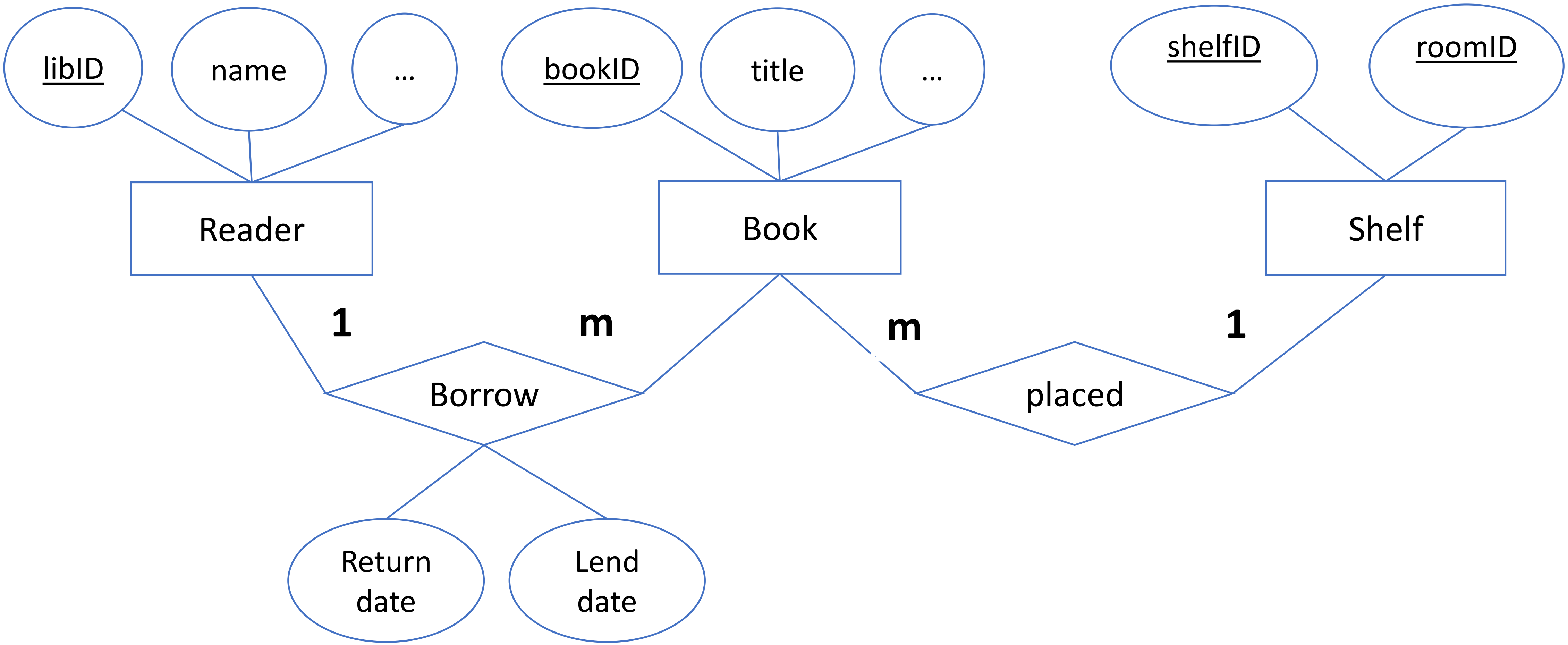
entity set E with
simple attribute A1,
composite attribute A2,
multivalued attribute A3,
derived attribute A4,
and primary key A1



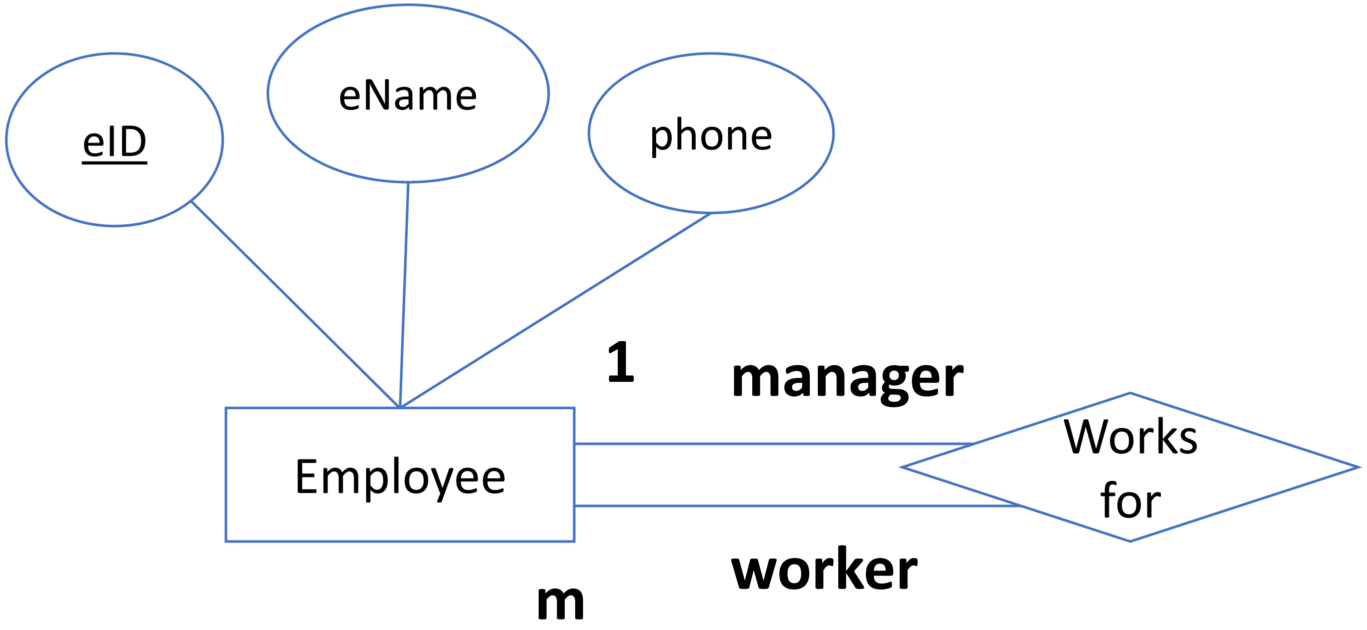
Alternative ER Notations



Example – cardinality



Example – **role**



Alternative ER Notations

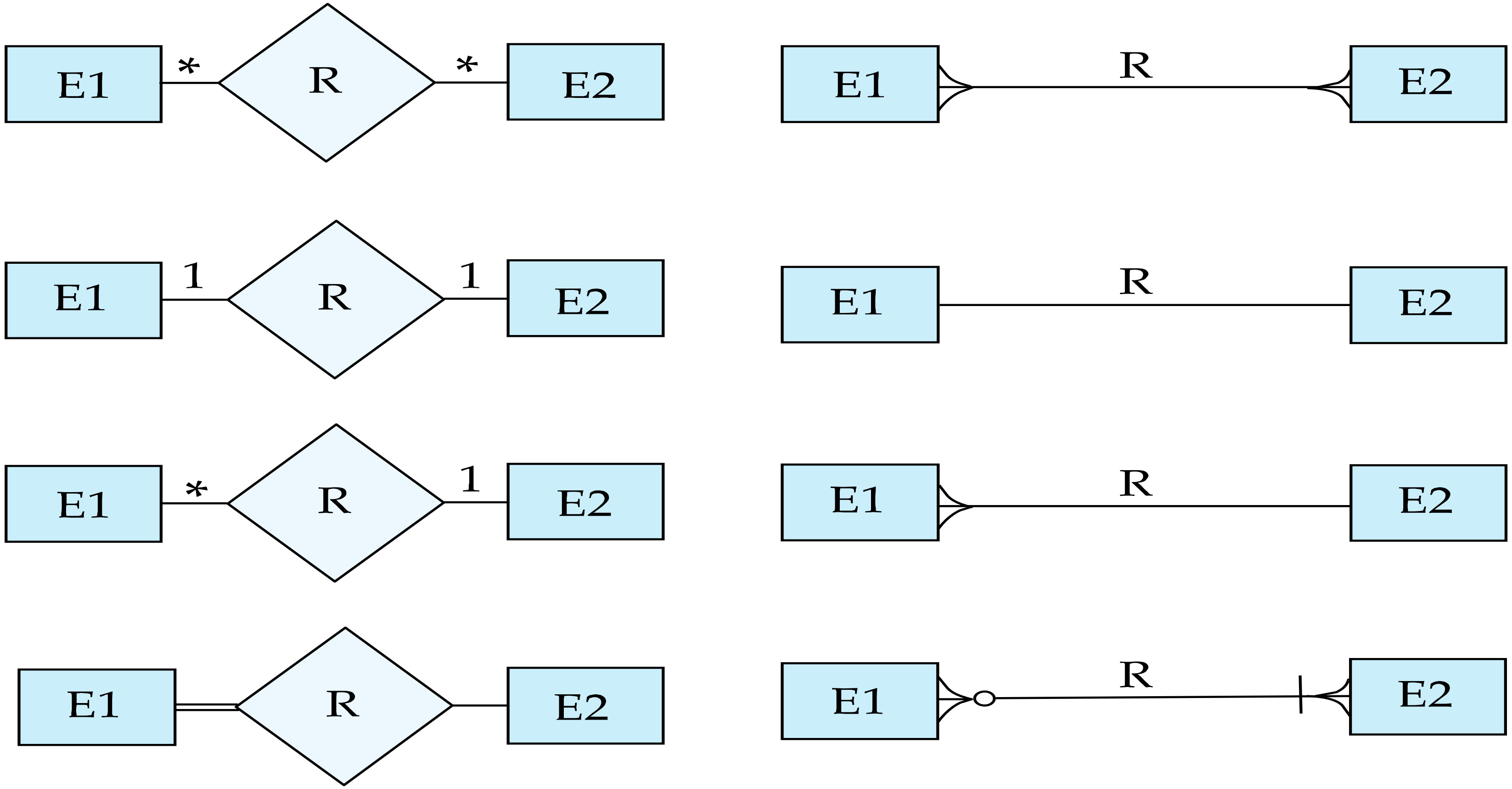
Chen | **IDE1FX (Crows feet notation)**

many-to-many relationship

one-to-one relationship

many-to-one relationship

participation in R: total (E1) and partial (E2)

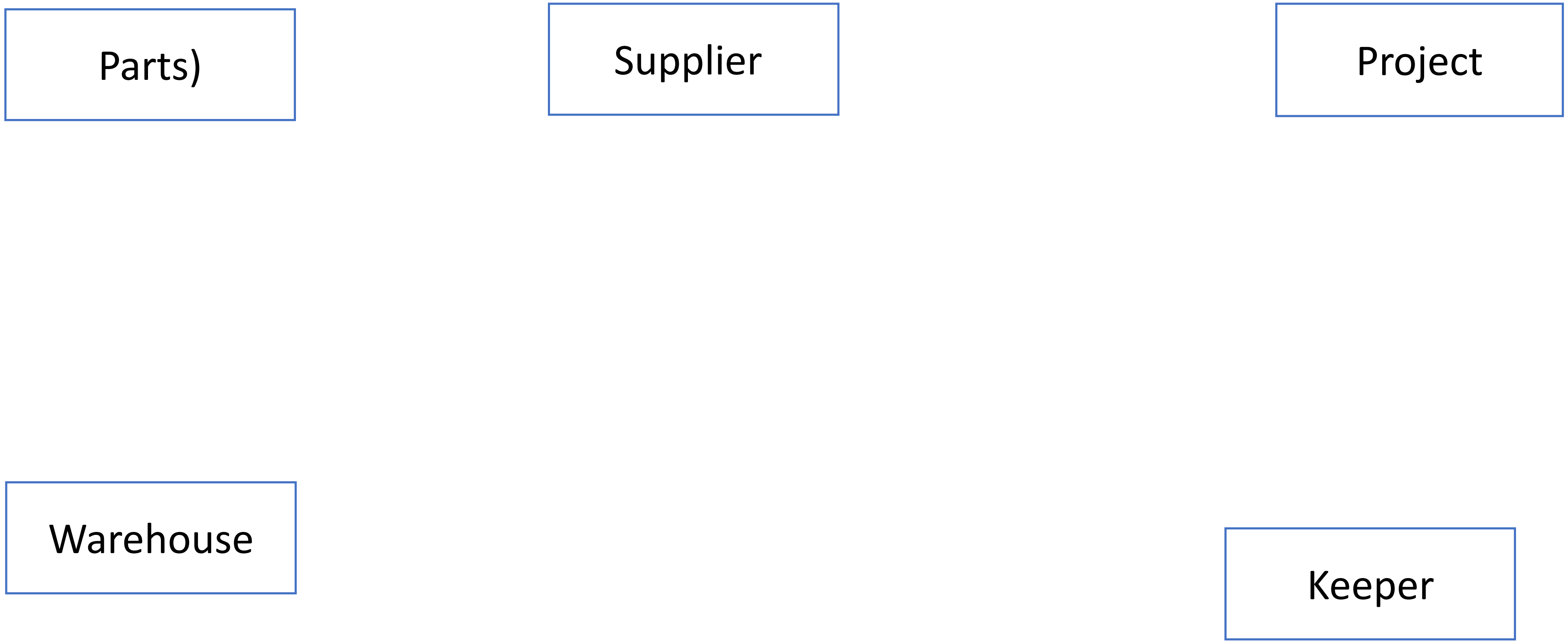


Case study – Inventory management

- Parts
- Supplier
  - Supply parts
- Project
  - Use parts
- Warehouse
  - Store parts
- Warehouse keeper
  - Manage warehouse
  - One keeper manages multiple warehouses

Step 1 – determine Entity sets

- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper



Parts)

Supplier

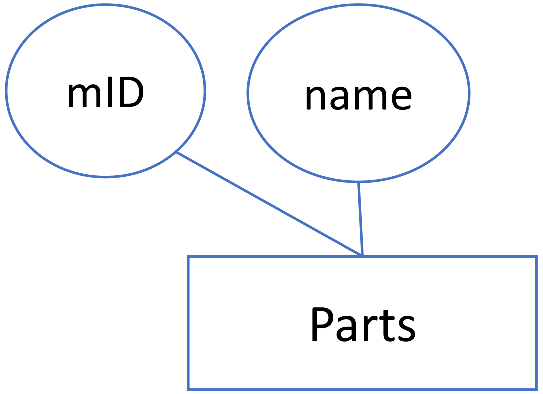
Project

Warehouse

Keeper

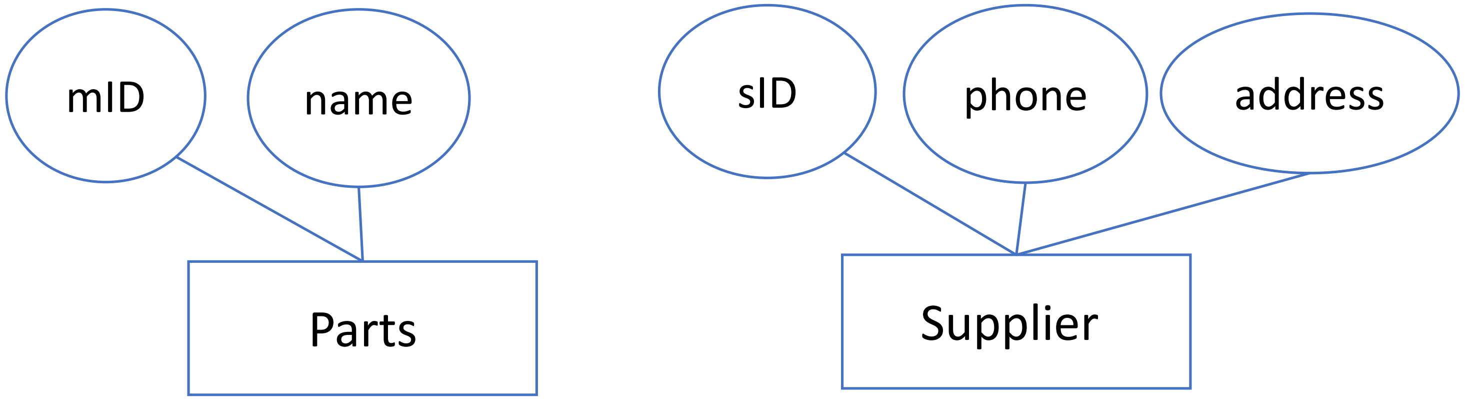
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 2 – determine Attributes



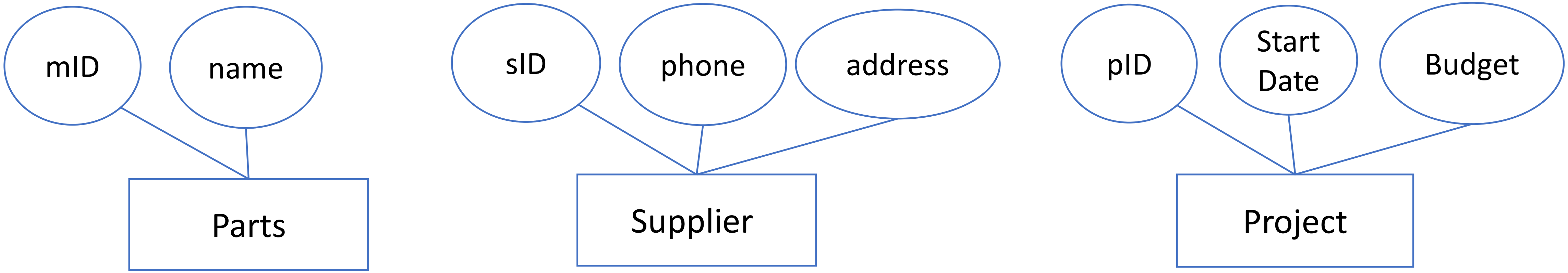
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 2 – determine Attributes



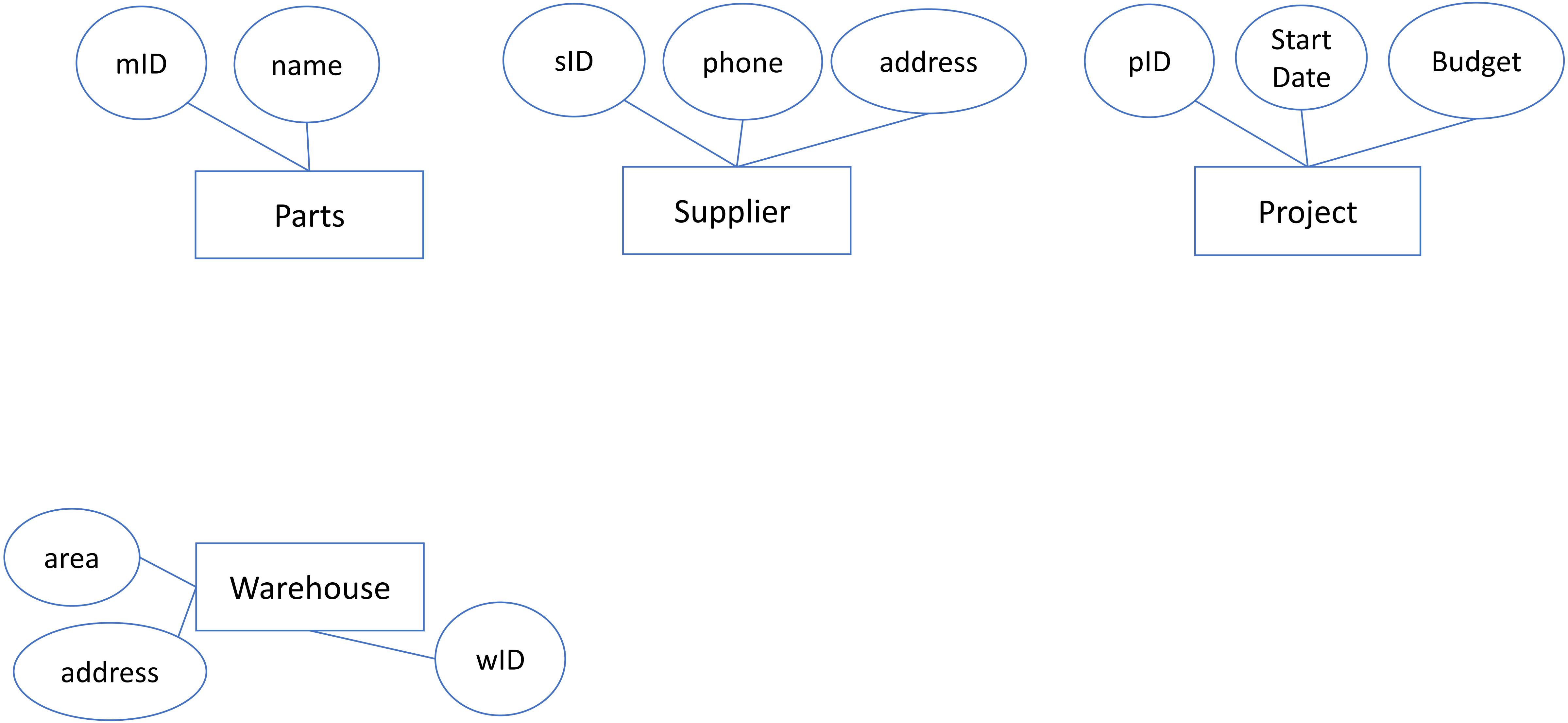
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 2 – determine Attributes



Step 2 – determine Attributes

- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper



mID

name

Parts

sID

phone

address

Supplier

pID

Start Date

Budget

Project

area

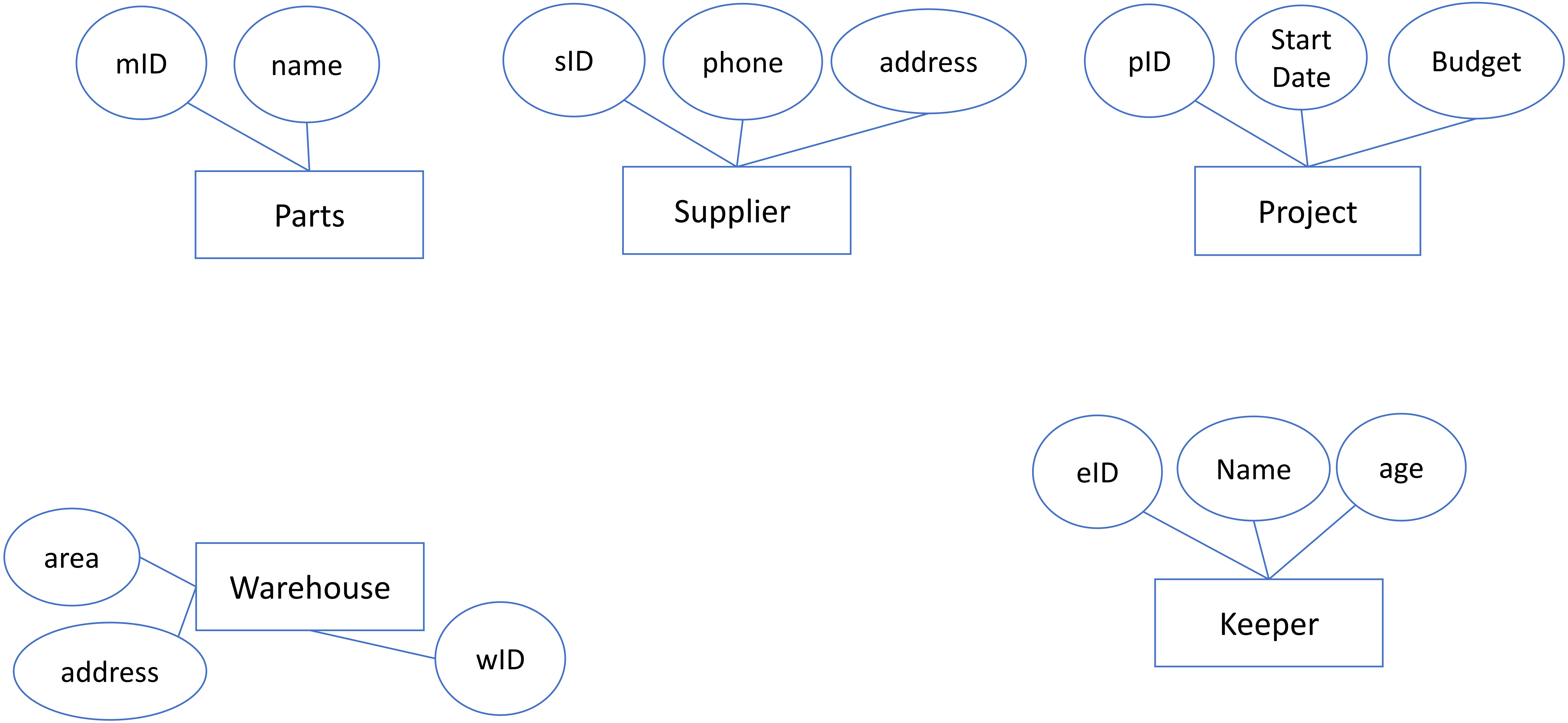
address

Warehouse

wID

Step 2 – determine Attributes

- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper



Parts: mID, name

Supplier: sID, phone, address

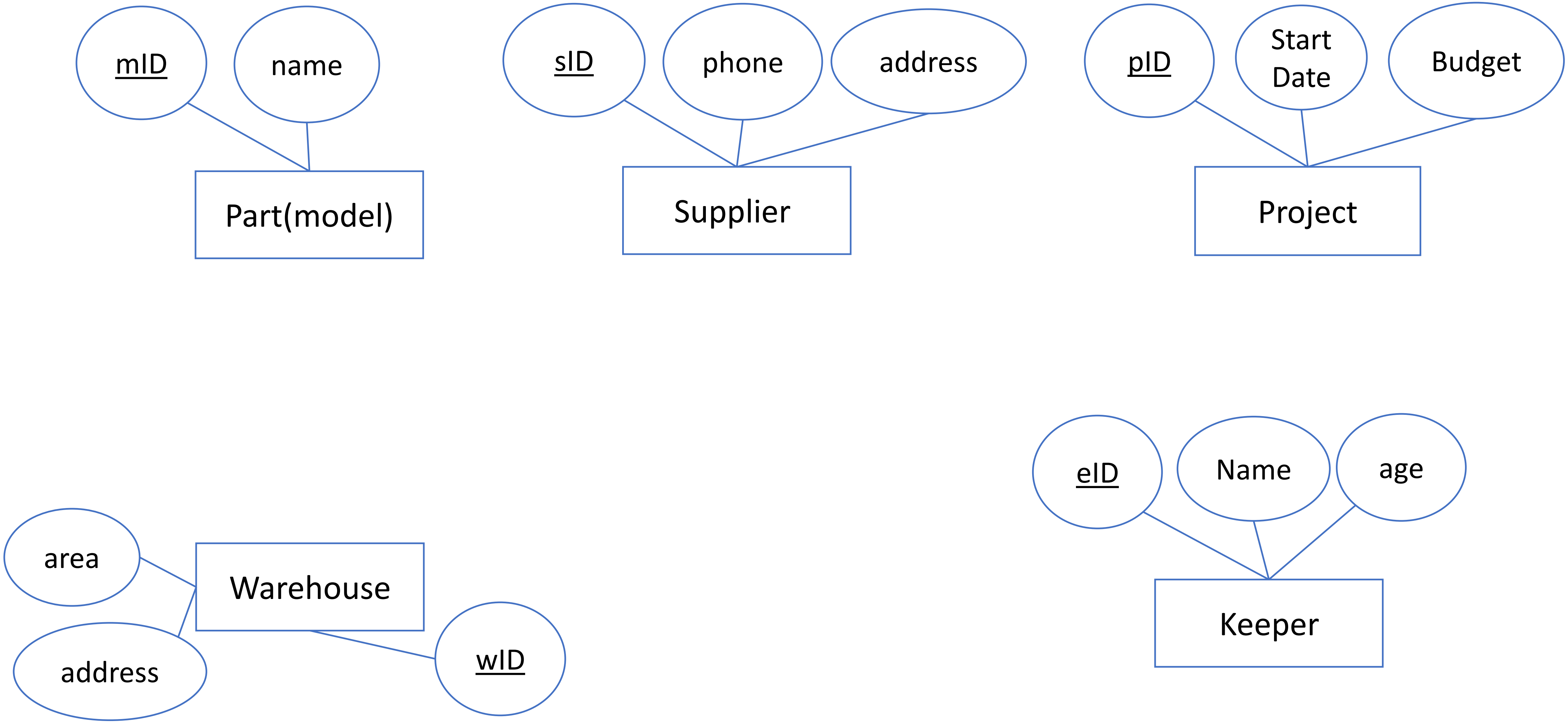
Project: pID, Start Date, Budget

Warehouse: area, address, wID

Keeper: eID, Name, age

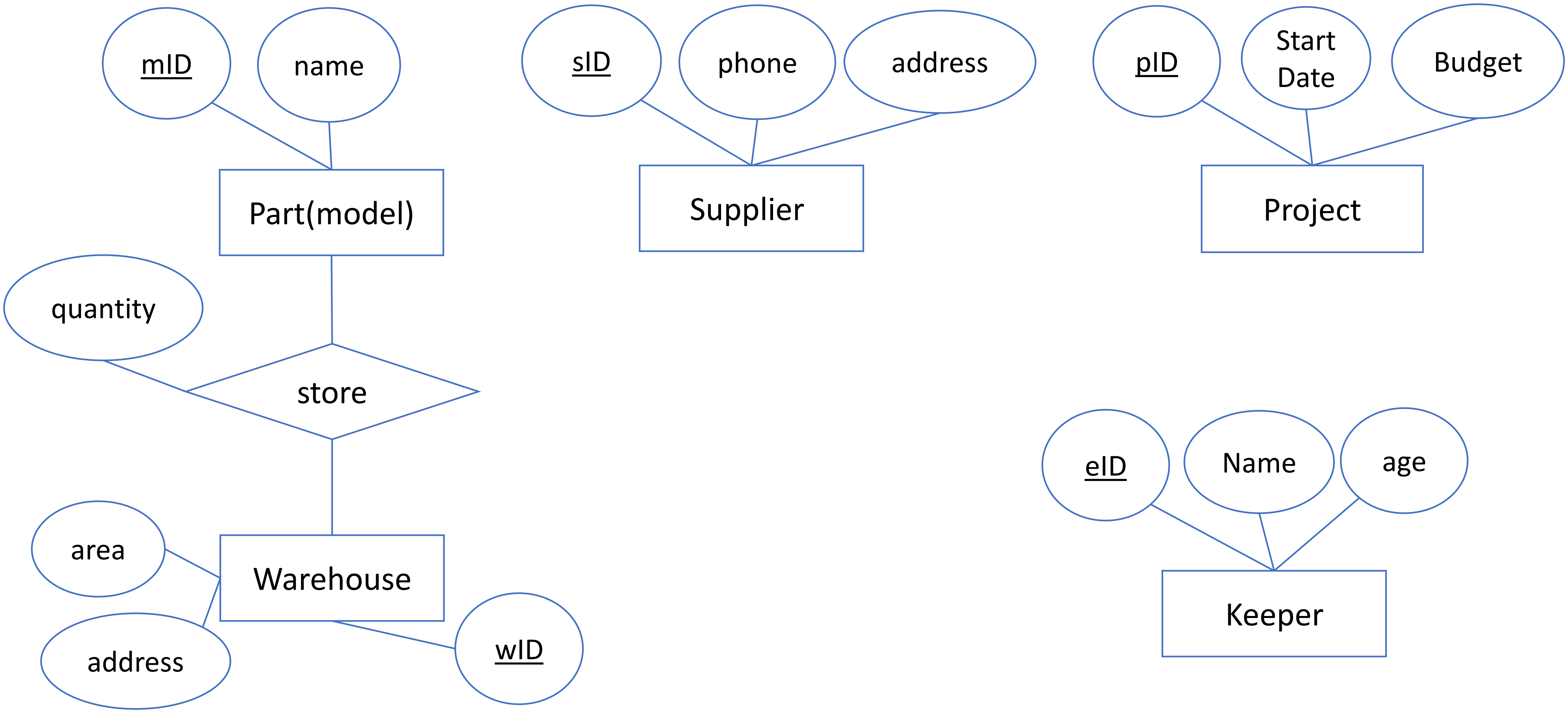
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 3 – determine keys



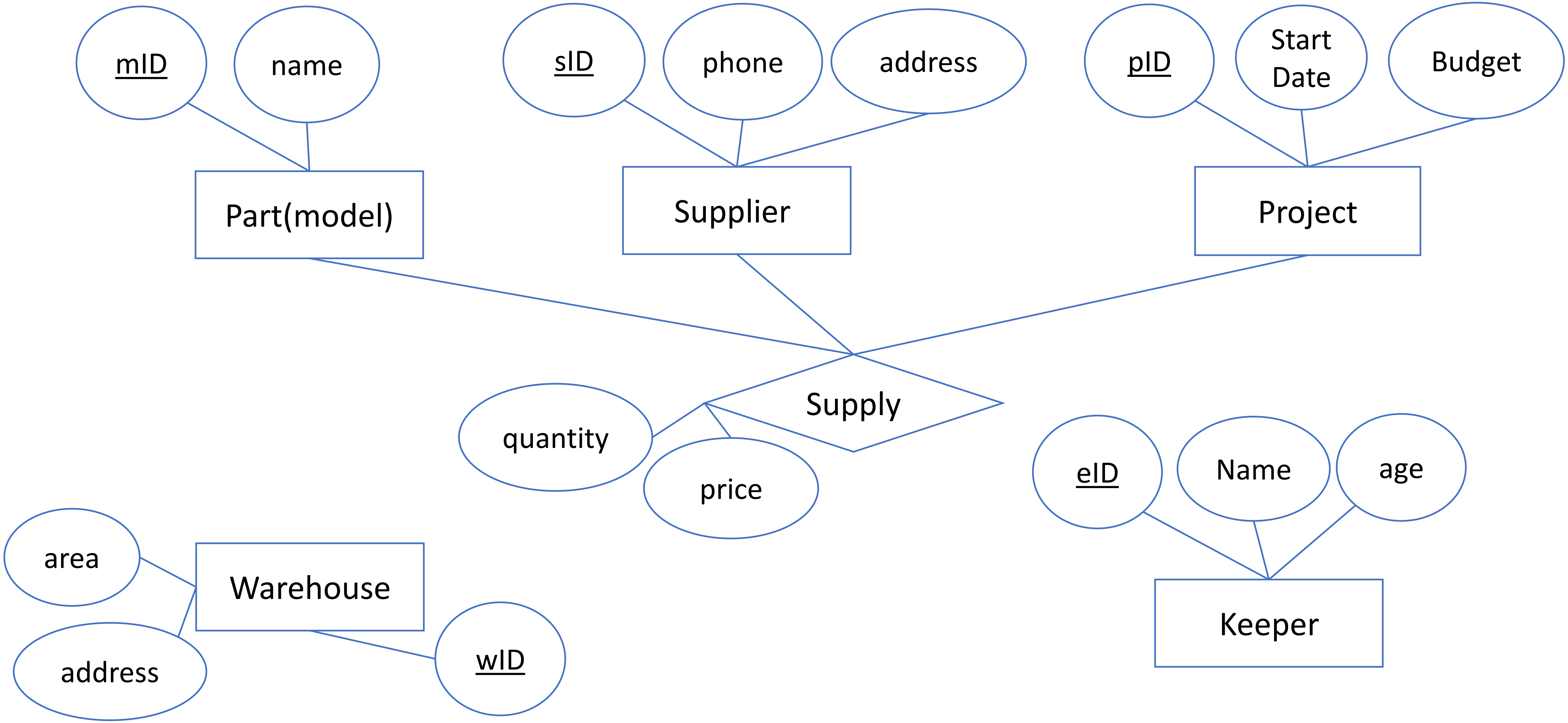
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 4 – establish relationships



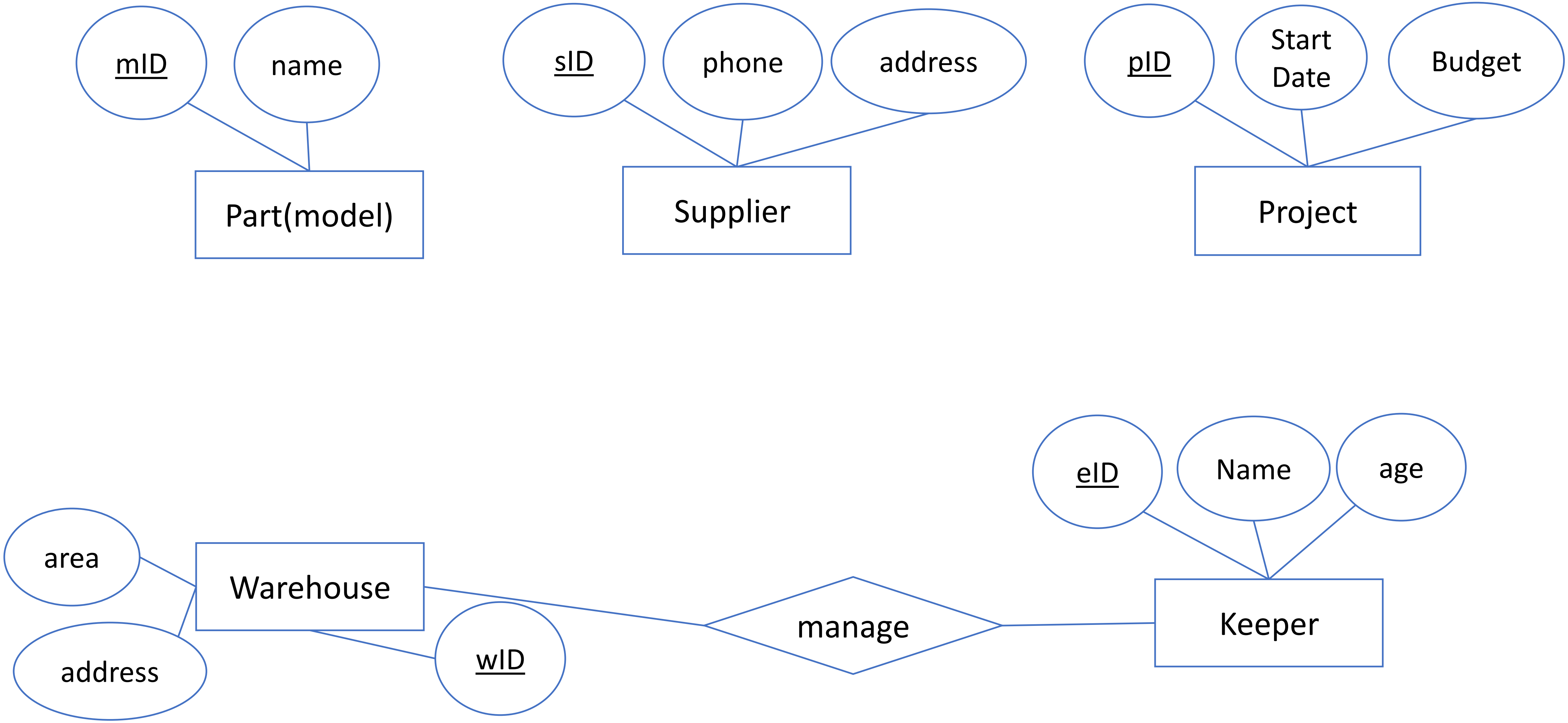
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 4 – establish relationships



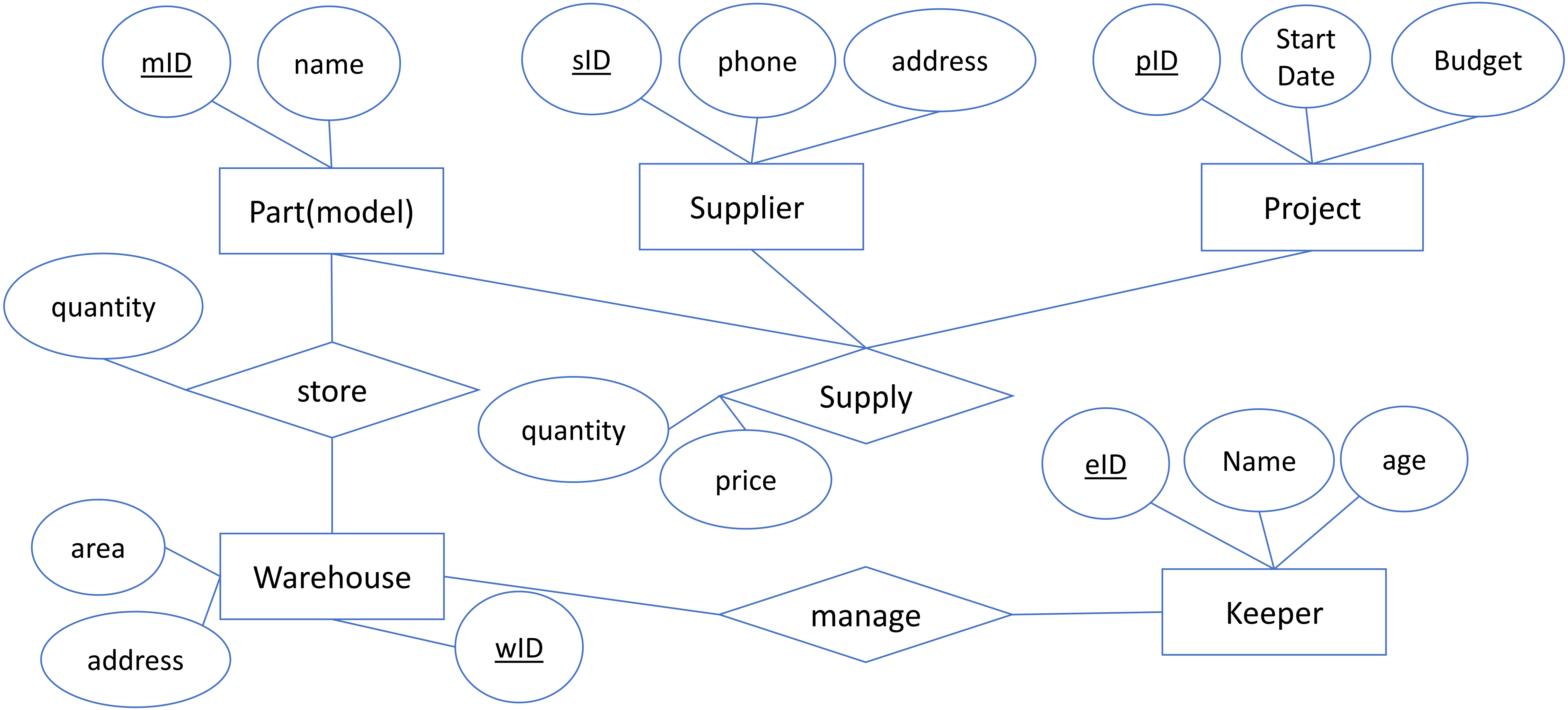
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 4 – establish relationships



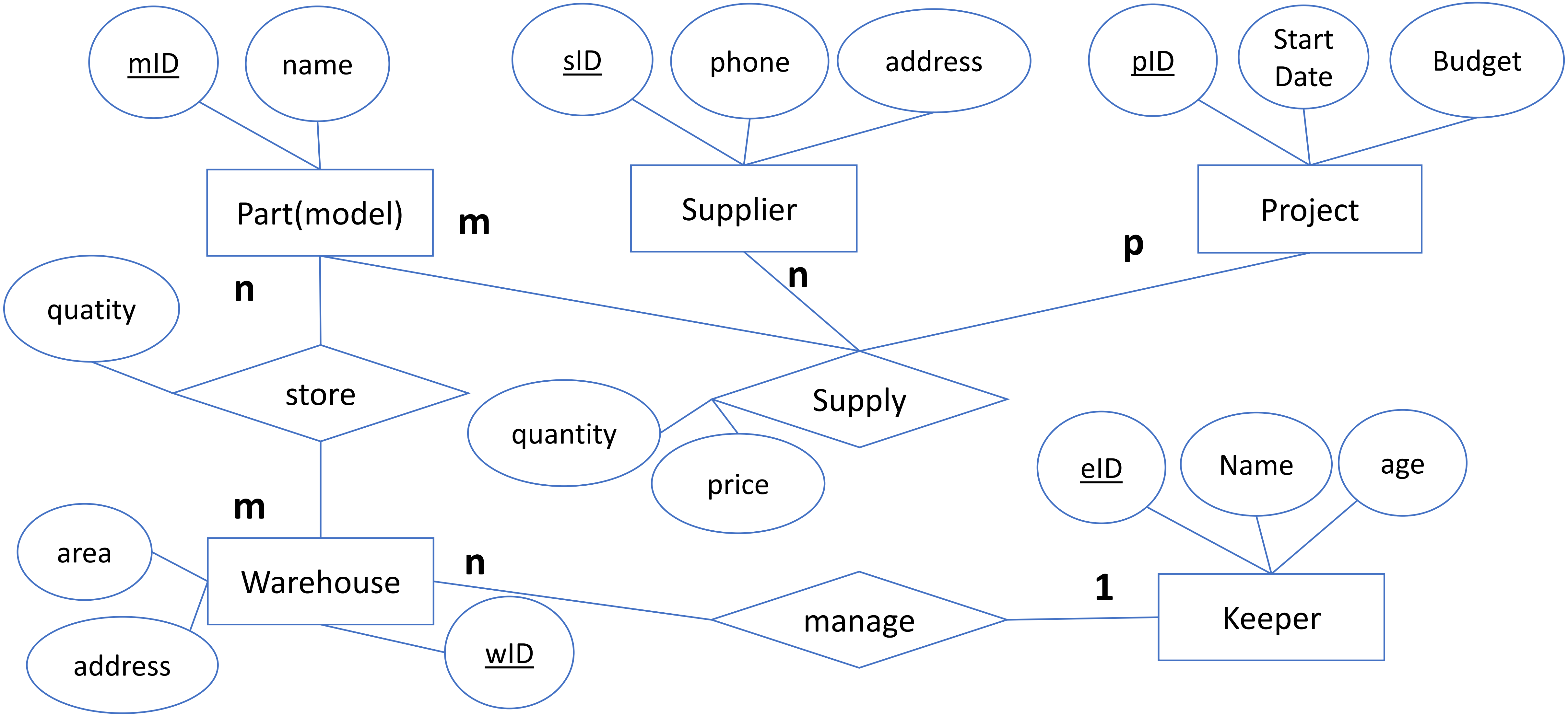
- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper

Step 4 – establish relationships



Step 5 – determine cardinality

- Parts
- Supplier
- Project
- Warehouse
- Warehouse keeper



Database for TV Companies (SELF STUDY)

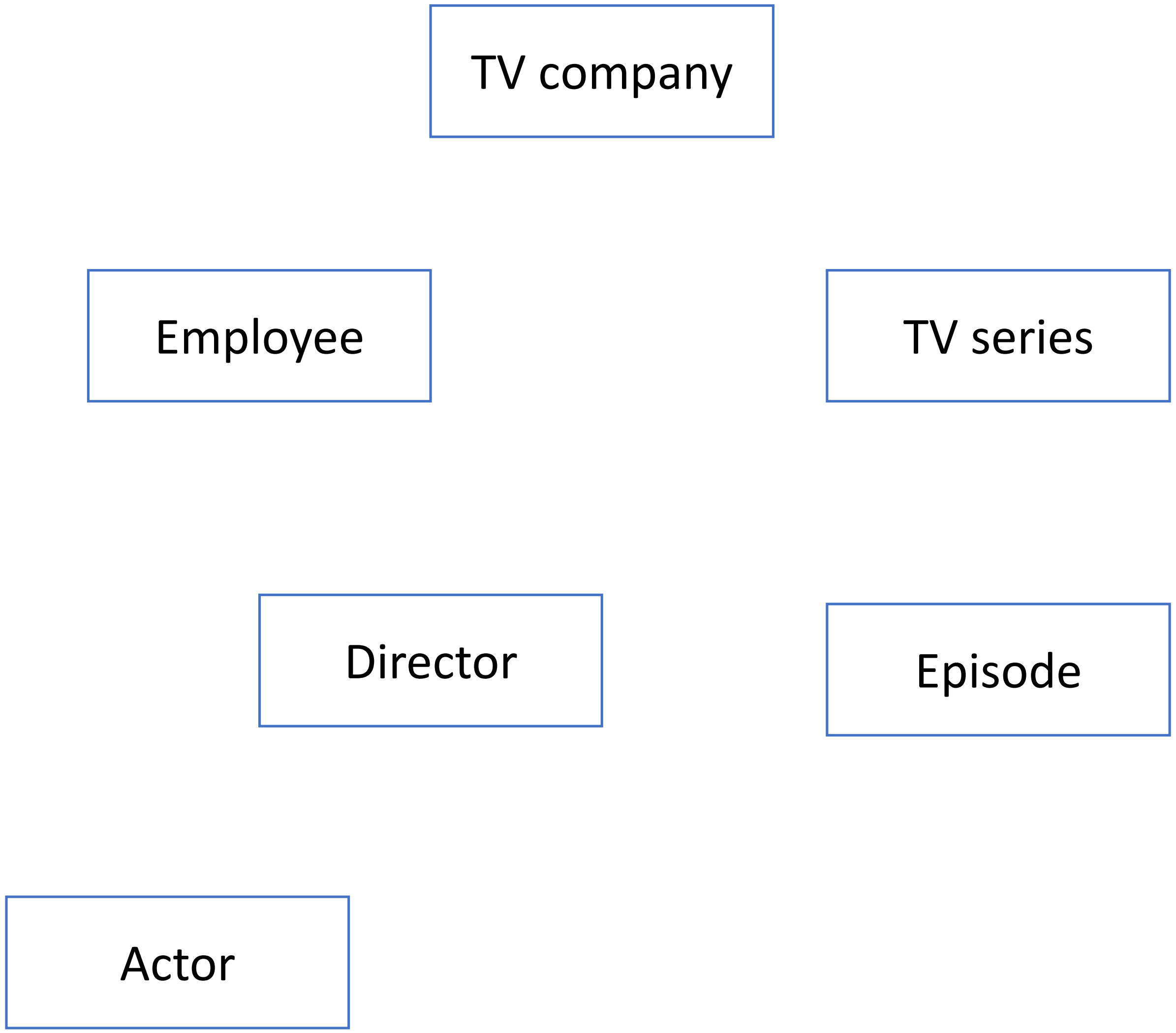
- Store data about the TV series that the company produces
- Data include information about actors who play in the series, and directors who direct the episodes of the series
  - Actors and directors are employed by the company
  - A TV series are divided into episodes
  - Each episode may be transmitted at several occasions
  - An actor is hired to participate in a series, but may participate in many series
  - Each episode of a series is directed by one of the directors, but different episodes may be directed by different directors.

Examples of database queries

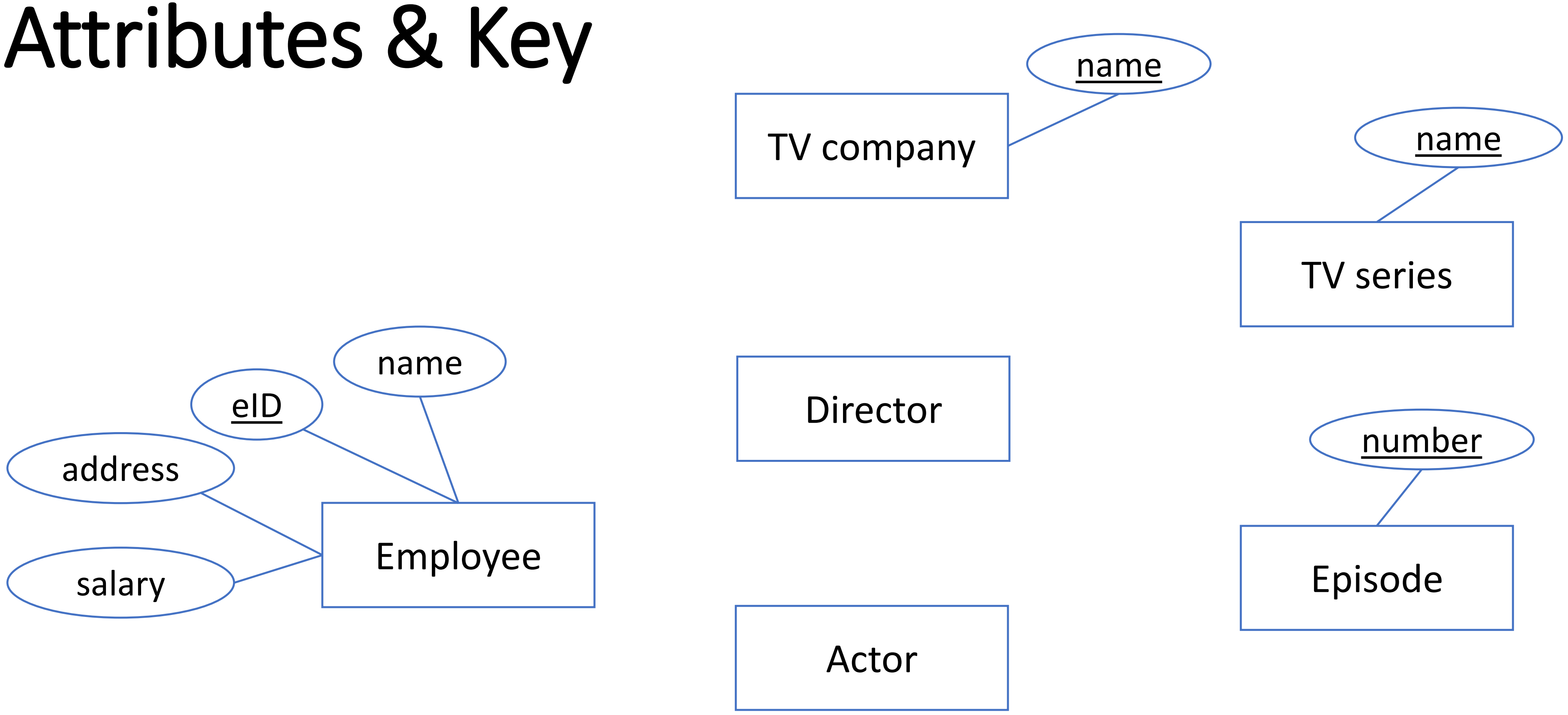
- Which actors play in the series The Night Of?
- In which series does the actor Carey Mulligan participate?
- Which actors participate in more than one series?
- How many times has the first episode of the series Wild Lies been transmitted? At what times?
- How many directors are employed by the company?
- Which directors has directed the greatest number of episodes?

Entity sets

- TV company
- TV series
- Episode
- Employee
- Actor
- Director



Attributes & Key



name

TV company

name

TV series

Director

number

Episode

Actor

name

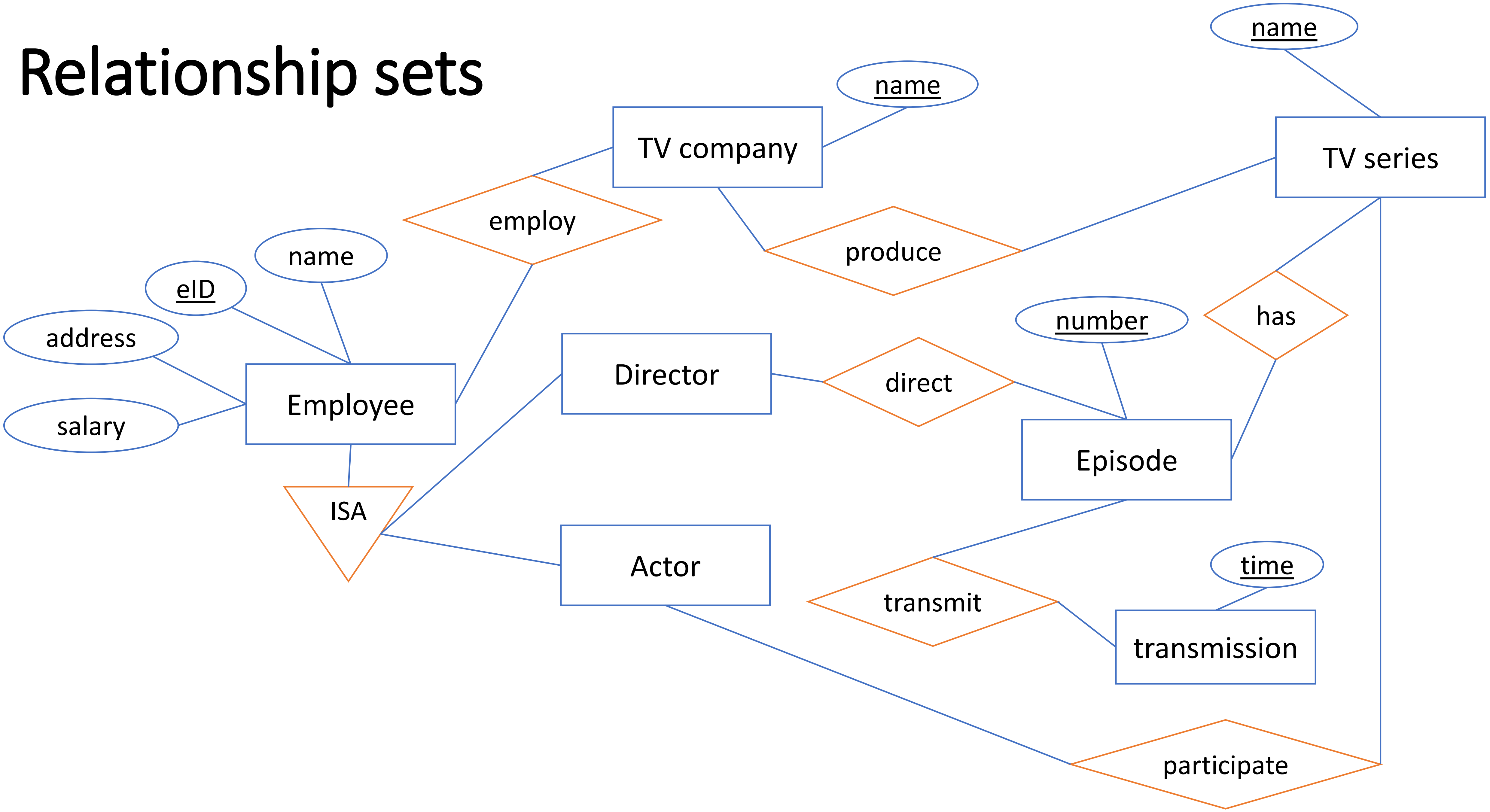
eID

address

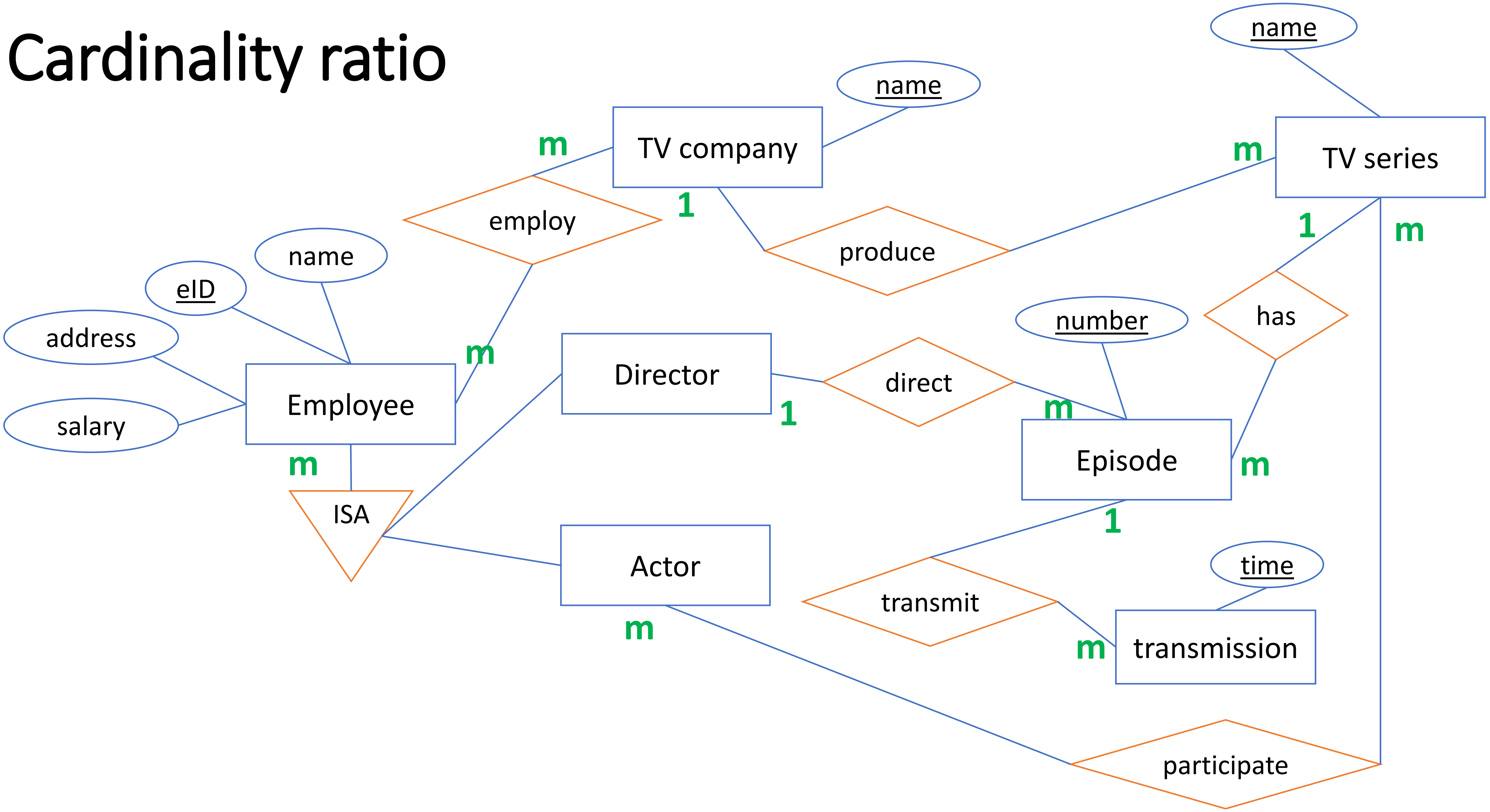
Employee

salary

Relationship sets



Cardinality ratio


name
name
TV company
TV series
m
m
employ
1
produce
1
m
name
eID
address
salary
Employee
m
has
number
Director
direct
1
m
Episode
m
m
ISA
1
Actor
m
time
transmit
transmission
m
participate