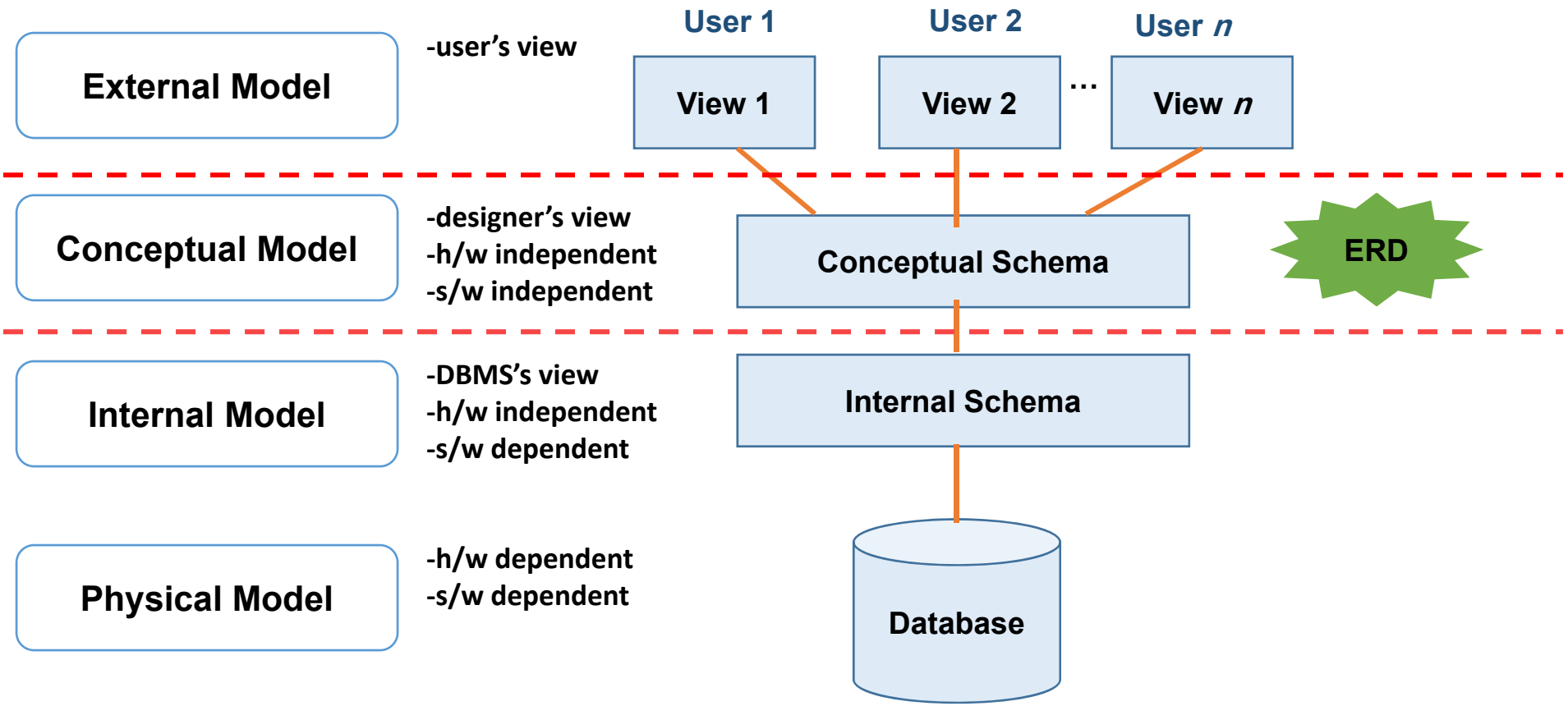
Database Design

Building ER diagram

Chapter 4: Entity Relationship (E-R) Modeling

Basic Modeling Concept

Three Level ANSI-SPARC Architecture



Example of a Database

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	04	King
92	CS1310	Fall	04	Anderson
102	CS3320	Spring	05	Knuth
112	MATH2410	Fall	05	Chang
119	CS1310	Fall	05	Anderson
135	CS3380	Fall	05	Stone

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

Figure 1.2
A database that stores student and course information.

Database system environment

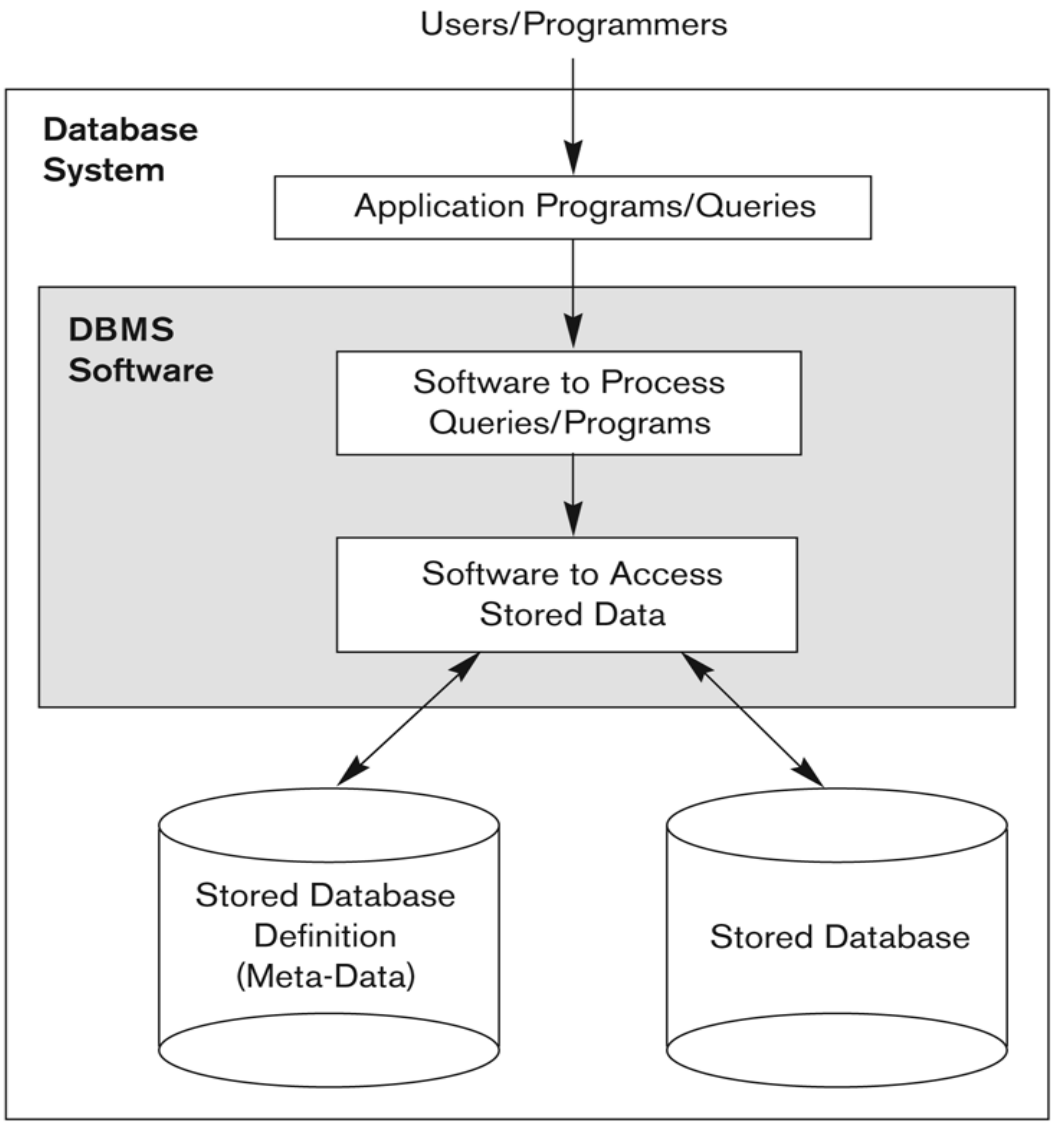


Figure 1.1
A simplified database system environment.

Main characteristics of the database approach

- Self-describing nature of a database system:
  - A DBMS catalog stores the description of a particular database (e.g. data structures, types, and constraints).
  - The description is called meta-data.
  - This allows the DBMS software to work with different database applications.
- Insulation between programs and data:
  - Called program-data independence.
  - Allows changing data structures and storage organization without having to change the DBMS access programs.
- Data abstraction:
  - A data model is used to hide storage details and present the users with a conceptual view of the database.
  - Programs refer to the data model constructs rather than data storage details.
- Support of multiple views of the data:
  - Each user may see a different view of the database, which describes only the data of interest to that user.

Database Design Process

- Two main activities:
  - Database design.
  - Applications design.
- Focus in this course on database design.
  - To design the conceptual, logical model for a database application.
- Applications design focuses on the programs and interfaces that access the database.
  - Generally considered part of software engineering.

Database Design Process

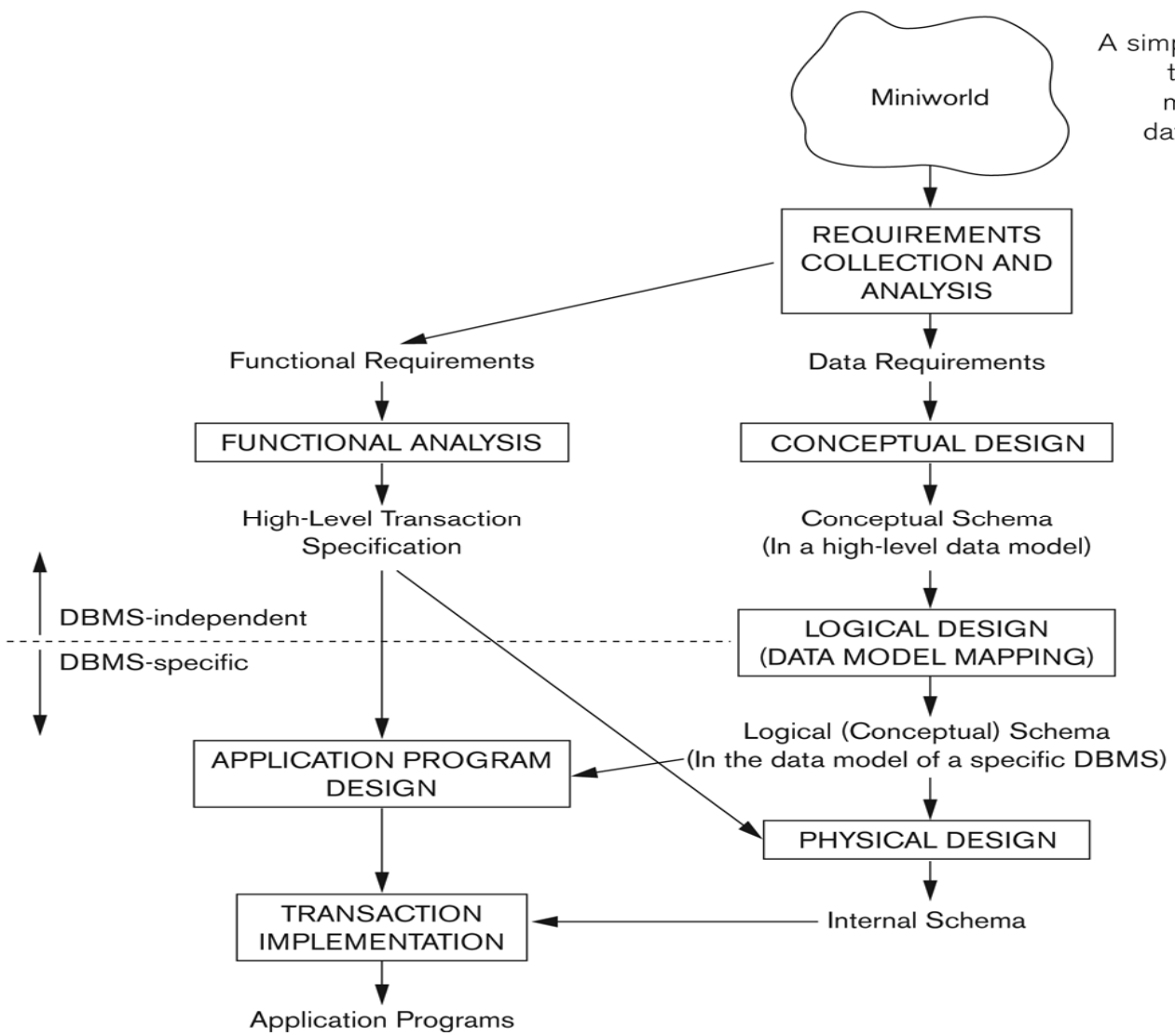
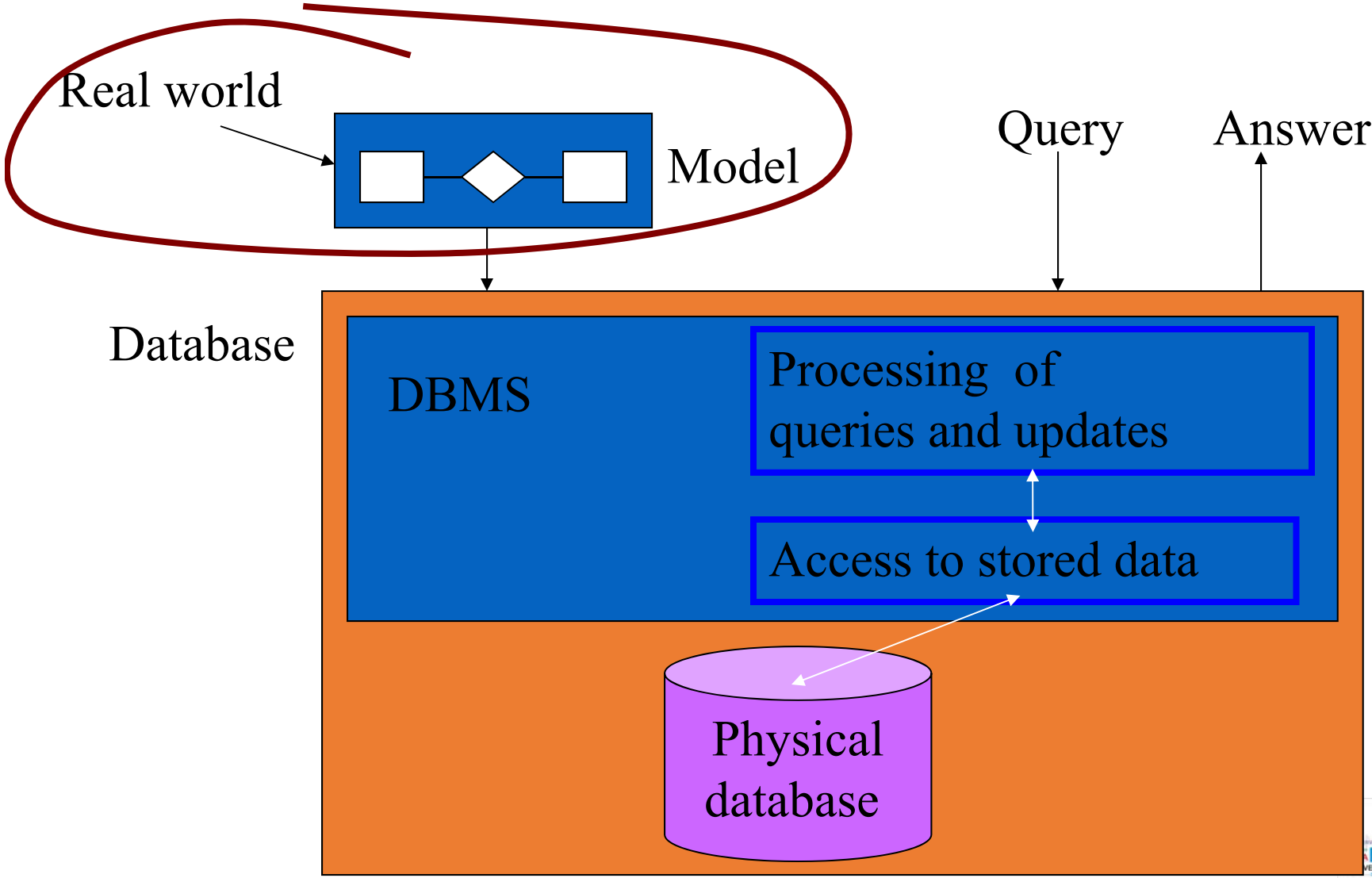


Figure 3.1
A simplified diagram to illustrate the main phases of database design.

Overview

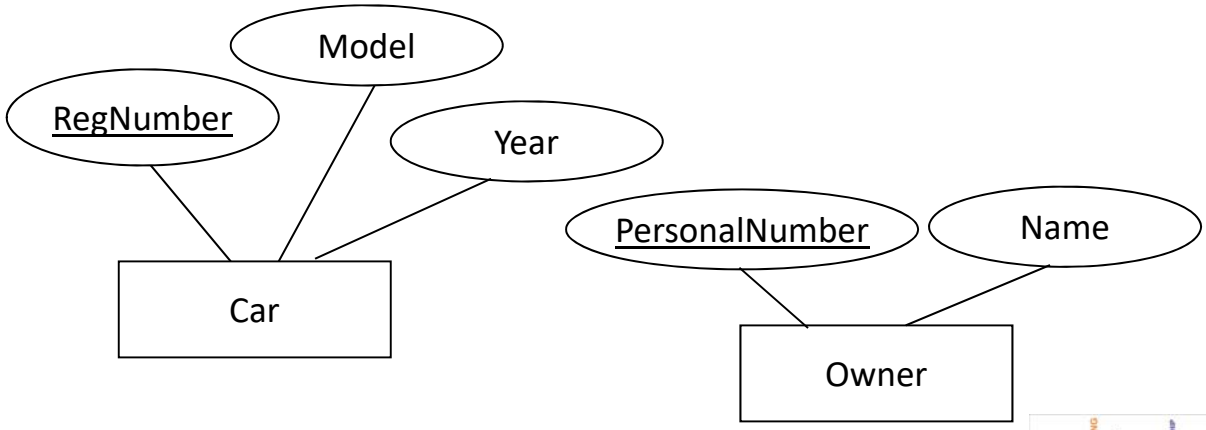

Real world
Model
Query
Answer
Database
DBMS
Processing of queries and updates
Access to stored data
Physical database

Entity-relationship (ER) model

- High-level conceptual data model.
  - An overview of the database.
  - Easy to discuss with non-database experts.
  - Easy to translate to data model of DBMS.
- ER diagram.
- Based on modelling objects in the real-world.

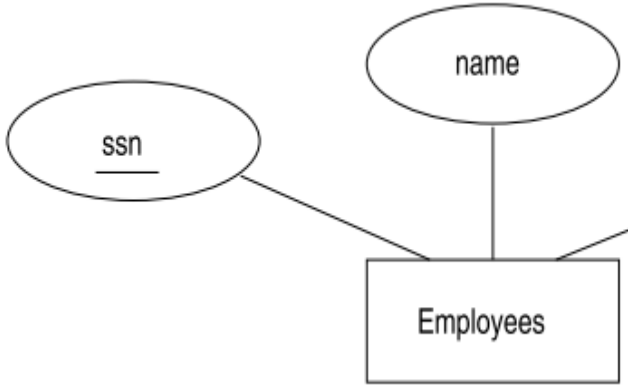
Entity and Entity type

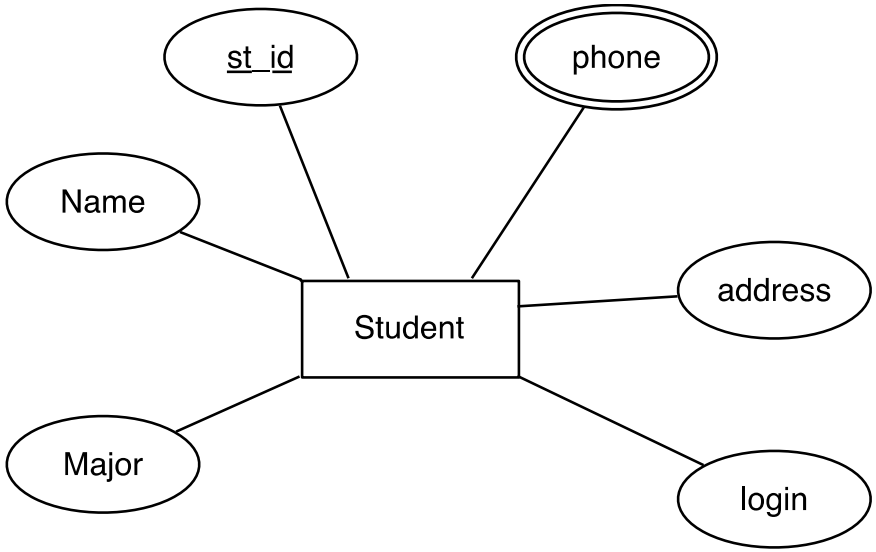
- Entity: A ”thing” in the real world with an independent existence.
- Attributes: Properties that describes an entity.
- Entity type: A collection of entities that have the same set of attributes.

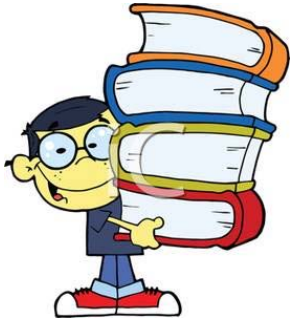


Conceptual Design

- Entities
- Attributes
- Attribute Domain
- Key
- Primary Key
- Candidate Keys

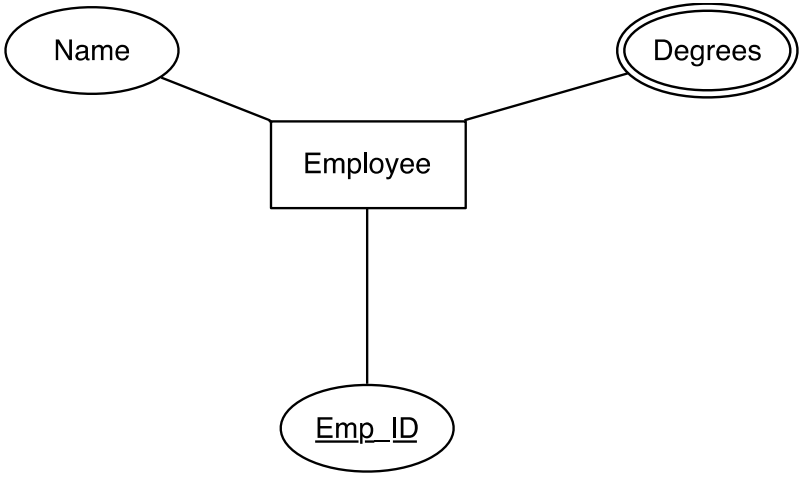






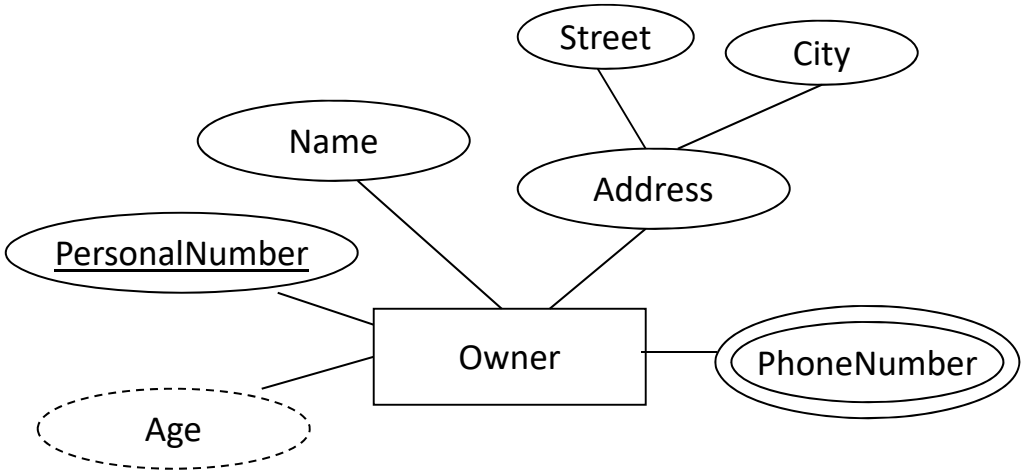
Multi-Valued Attributes

- Can have many values.
  - Examples:
    - A person may have several college degrees.
    - A household may have several phones with different numbers
    - A student has hobbies



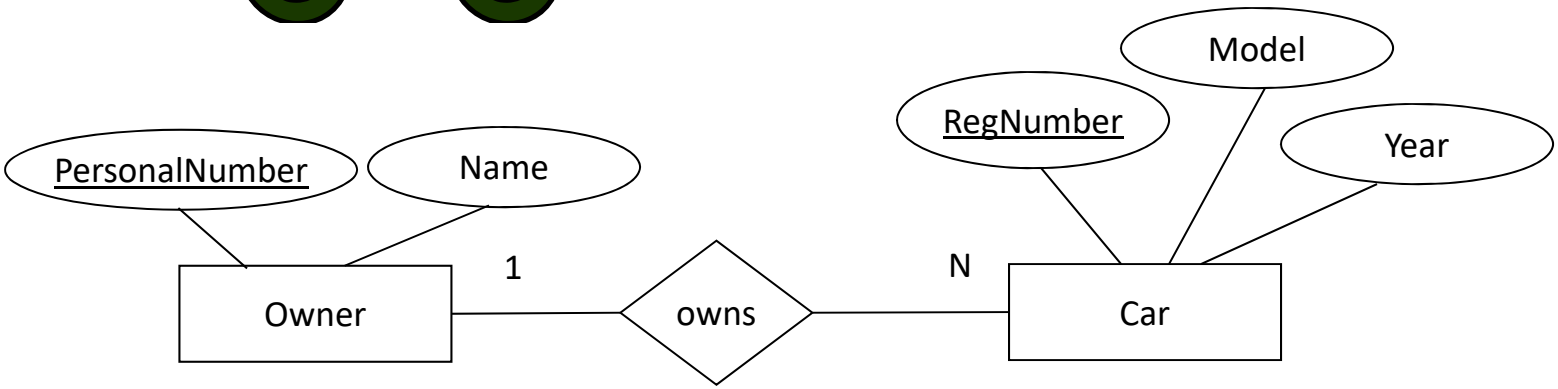
Constraints on Attributes

- Value sets (domains) of attributes.
- Key attributes.



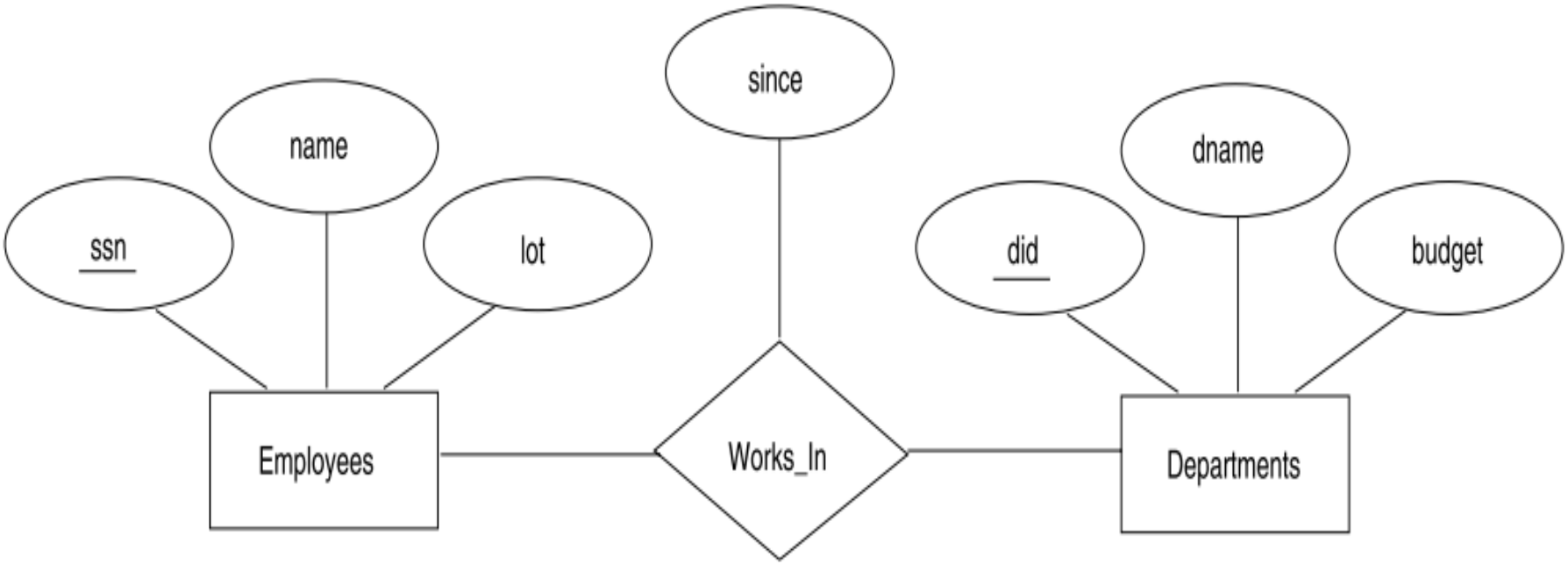
Relationship Type

- Relationship type: Association among entity types.

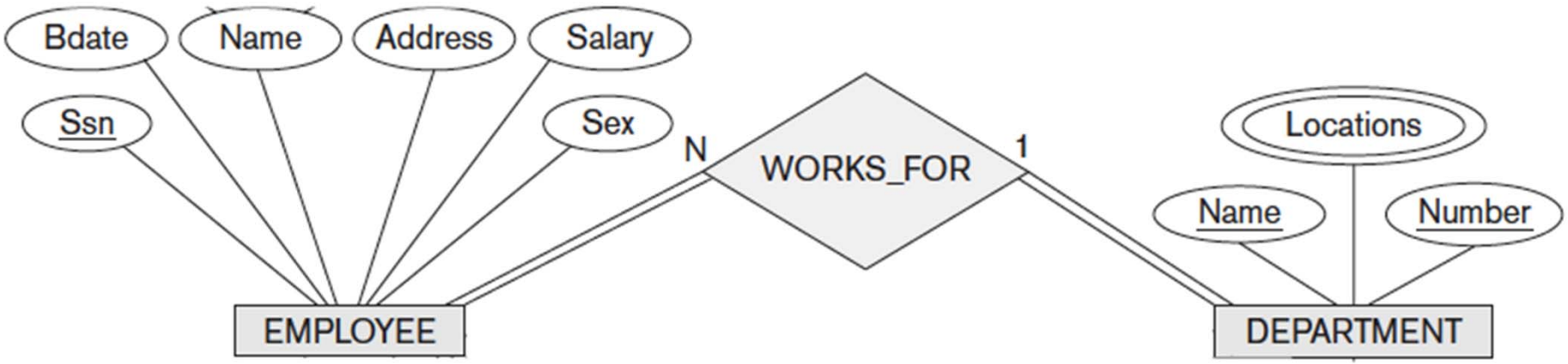


Conceptual Design 2

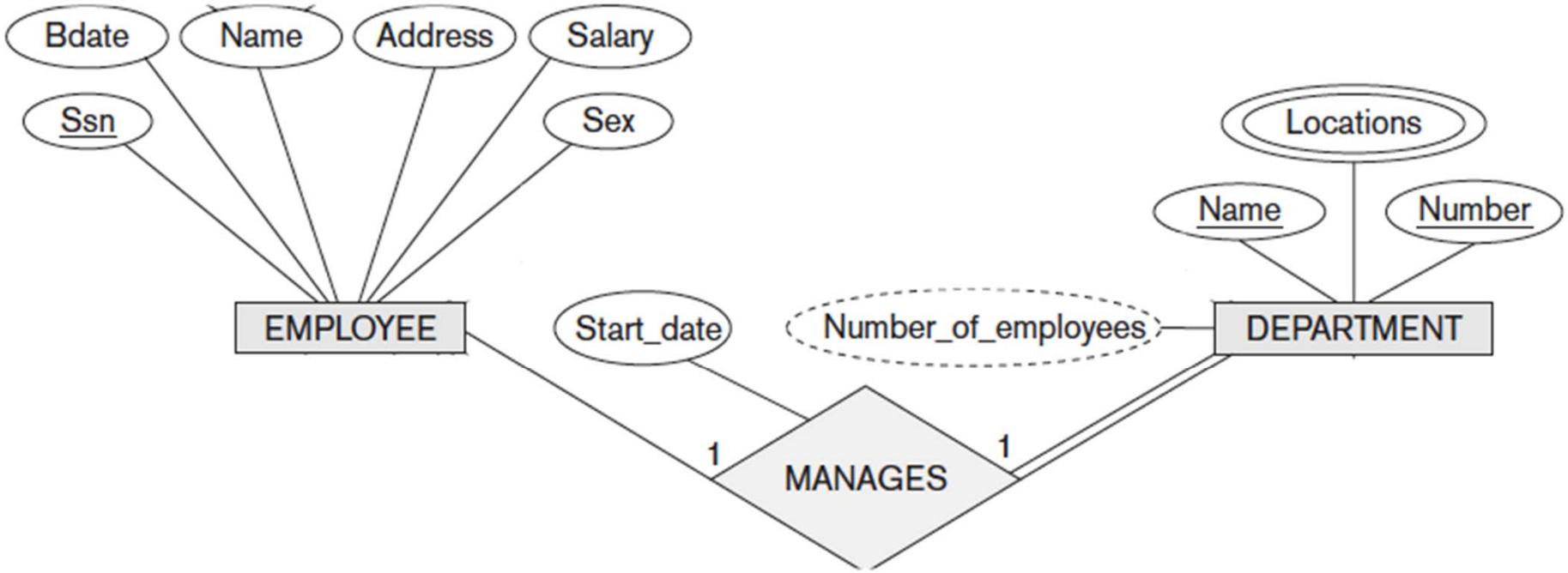
- Relationships
- Descriptive Attributes

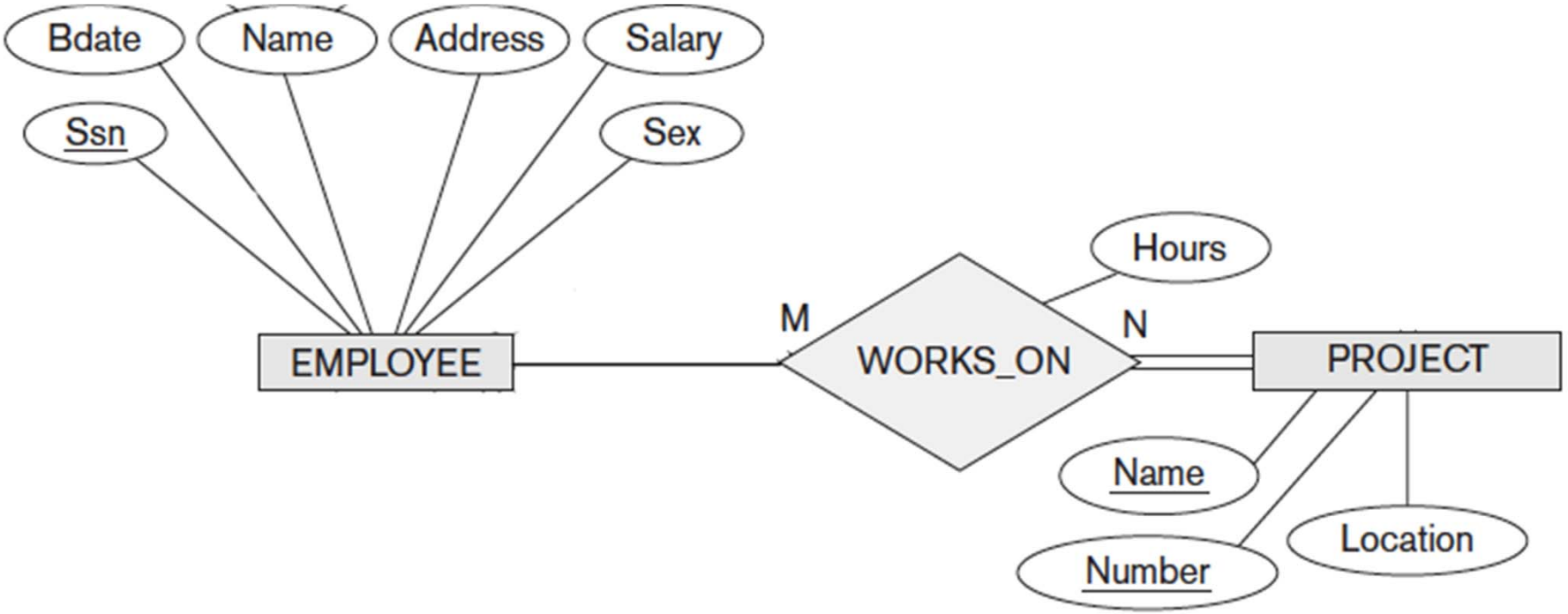


Examples: Employee



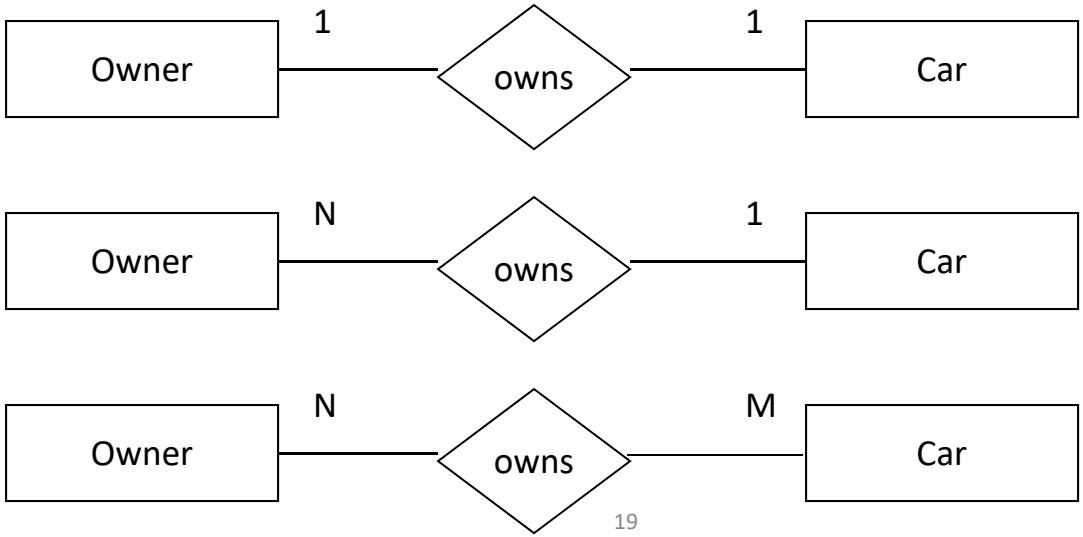
Examples: Manage



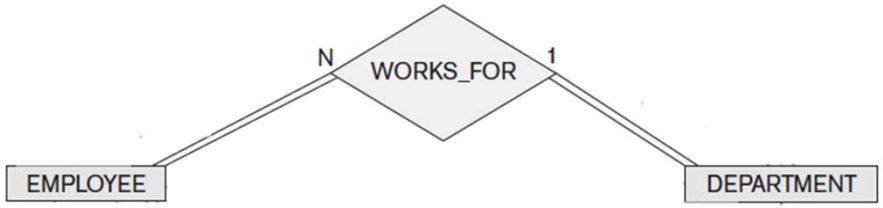

Bdate
Name
Address
Salary
Ssn
Sex
EMPLOYEE
M
WORKS_ON
N
Hours
PROJECT
Name
Location
Number

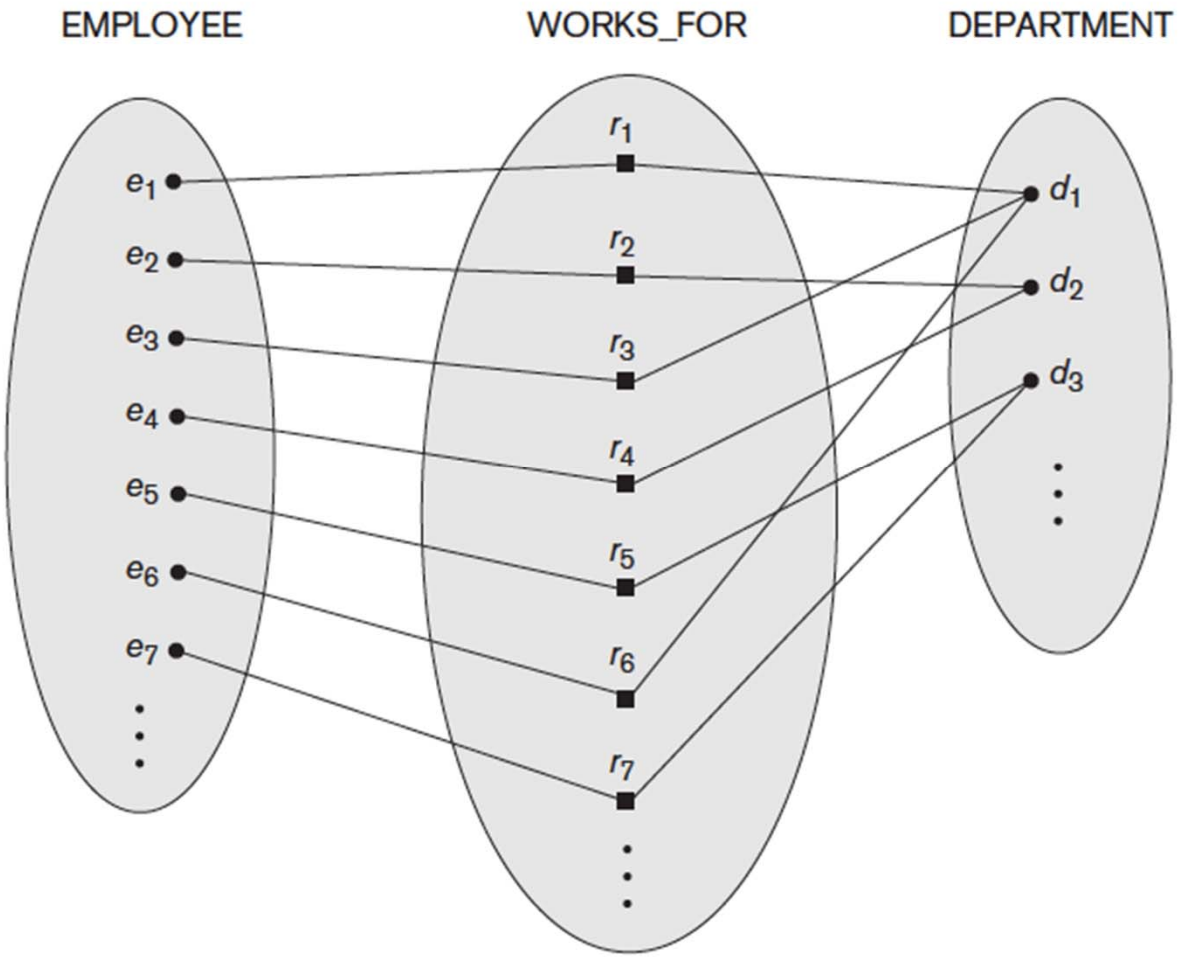
Constraints on relationship types

- Cardinality ratio: **Maximum** number of relationships an entity can participate in.
- Possible cardinality ratio: 1:1, 1: N, N:1, and N:M



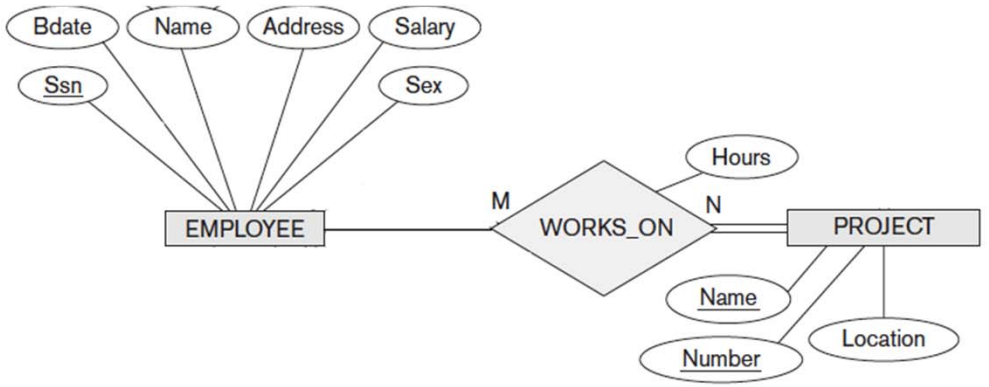
One-to-Many

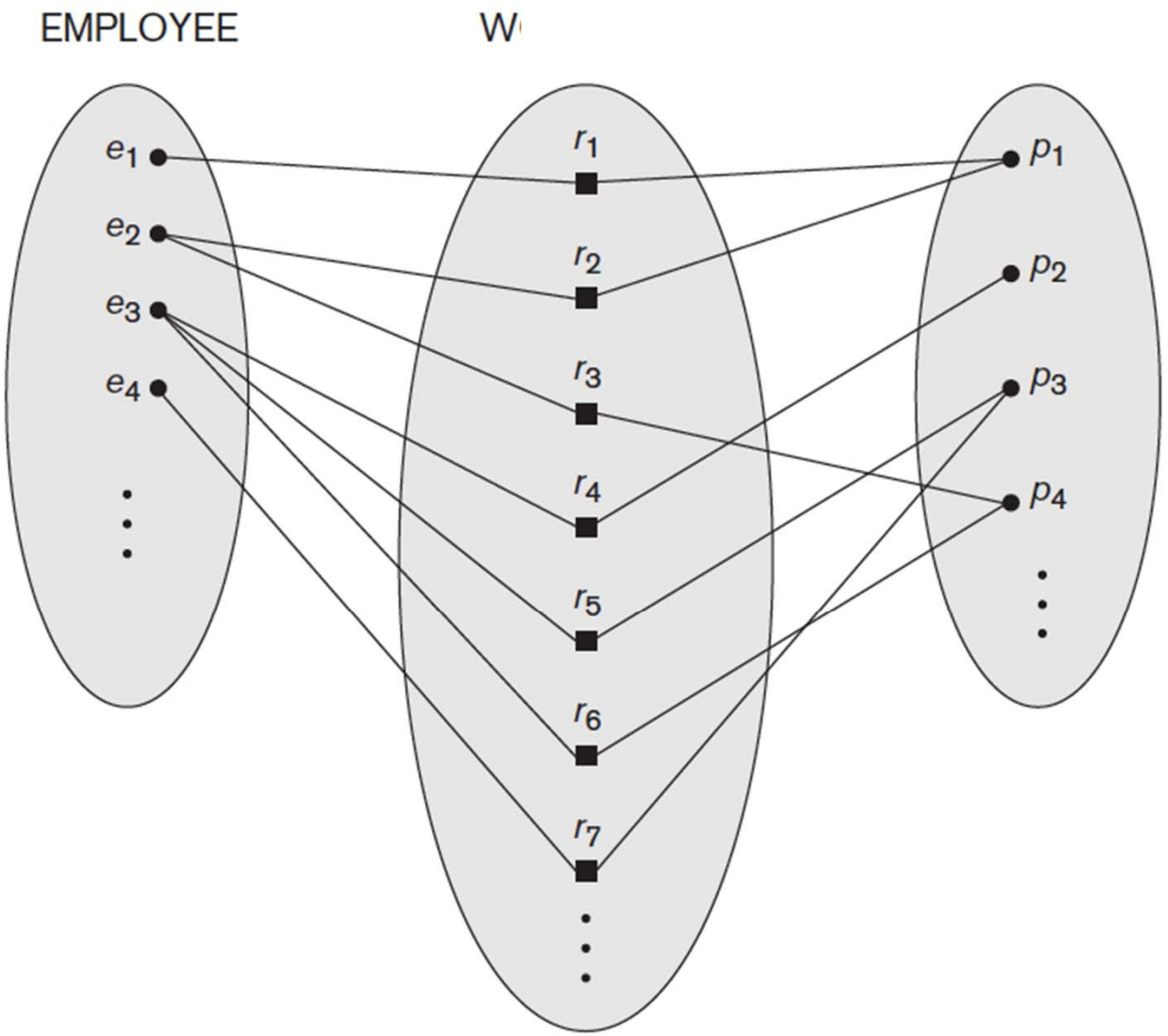

N
WORKS_FOR
1
EMPLOYEE
DEPARTMENT


EMPLOYEE
WORKS_FOR
DEPARTMENT
e_1
e_2
e_3
e_4
e_5
e_6
e_7
r_1
r_2
r_3
r_4
r_5
r_6
r_7
d_1
d_2
d_3


DATABASE
RELATIONAL
DBMS
DATA
DEVELOPER
SECURITY
KEY

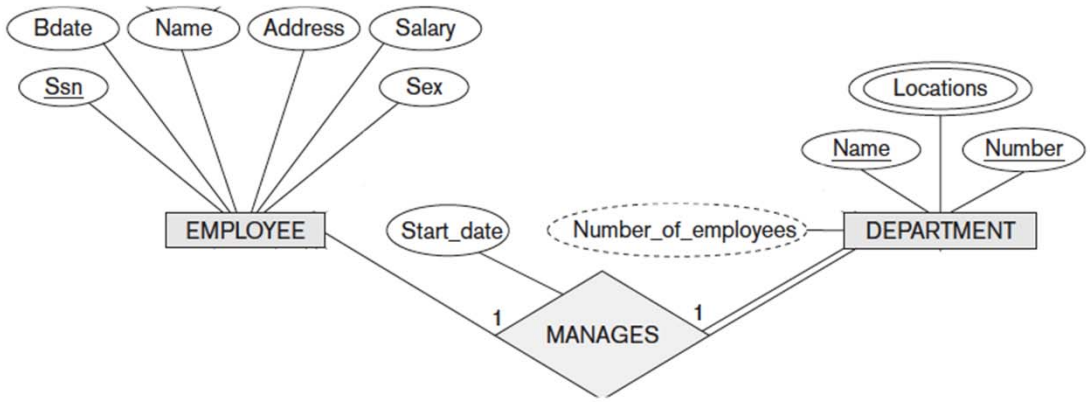
Many-to-Many

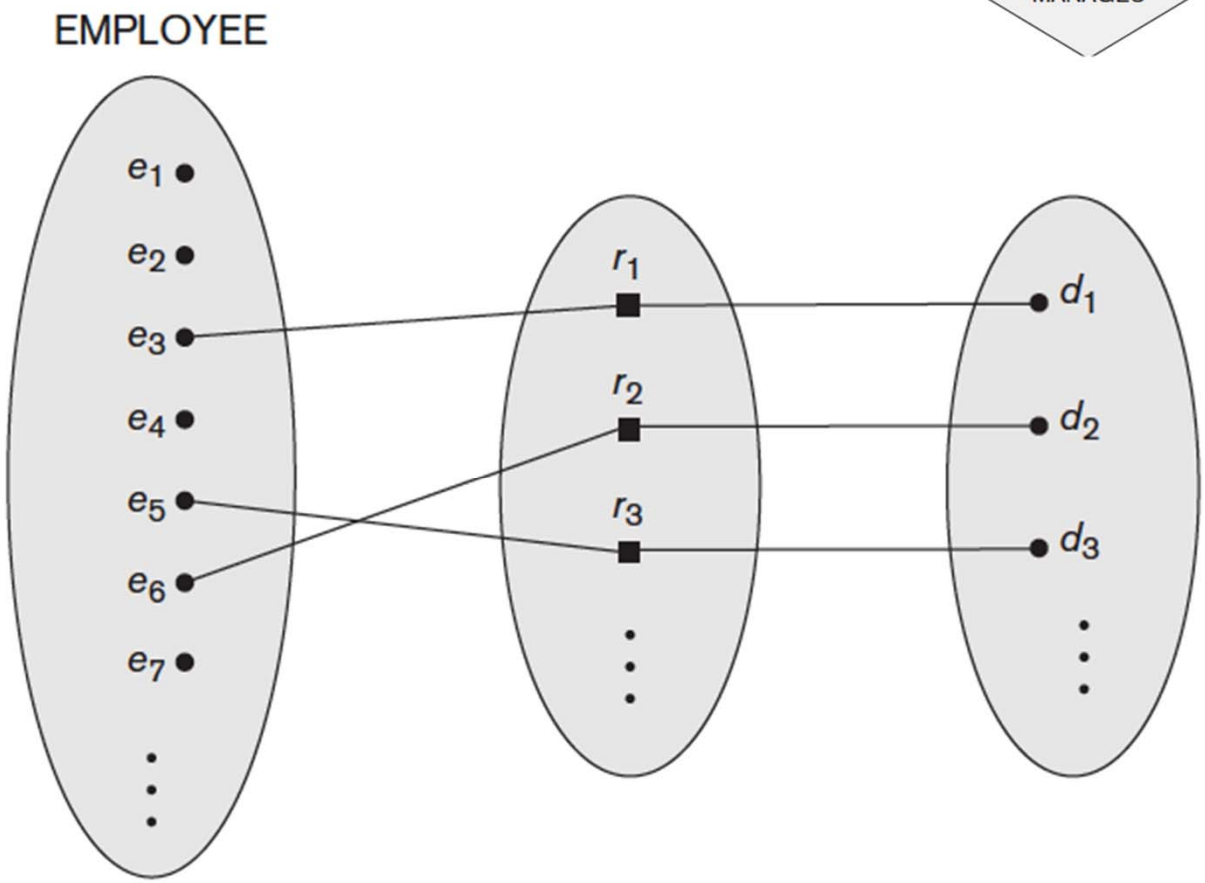

Bdate
Name
Address
Salary
Ssn
Sex
EMPLOYEE
M
WORKS_ON
N
Hours
PROJECT
Name
Number
Location


EMPLOYEE
W
e_1
e_2
e_3
e_4
r_1
r_2
r_3
r_4
r_5
r_6
r_7
p_1
p_2
p_3
p_4


DATABASE
DATA
RELATIONAL
DESIGNER
DBMS
DEVELOPER
SECURITY
KEY
MIRRORING

One-to-One


Bdate
Name
Address
Salary
Ssn
Sex
Locations
Name
Number
EMPLOYEE
Start_date
Number_of_employees
DEPARTMENT
1
MANAGES
1


EMPLOYEE
e_1
e_2
e_3
e_4
e_5
e_6
e_7
r_1
r_2
r_3
d_1
d_2
d_3


DATABASE
RELATIONAL
DBMS
DATA
DEVELOPER
SECURITY

Keys

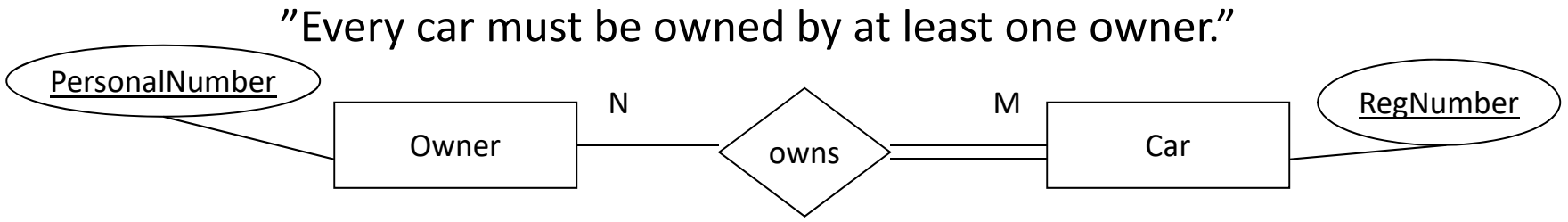
- A **super key** of an entity set is a set of one or more attributes whose values uniquely determine each entity.
- A candidate key of an entity set is a minimal super key
  - ID is **candidate key** of instructor
  - course_id is candidate key of course
- Although several candidate keys may exist, one of the candidate keys is selected to be the **primary key.**

Keys for Relationship Sets

- The combination of primary keys of the participating entity sets forms a super key of a relationship set.
  - (s_id, i_id) is the super key of advisor
  - NOTE: this means a pair of entity sets can have at most one relationship in a particular relationship set.
    - Example: if we wish to track multiple meeting dates between a student and her advisor, we cannot assume a relationship for each meeting. We can use a multivalued attribute though
- Must consider the mapping cardinality of the relationship set when deciding what are the candidate keys
- Need to consider semantics of relationship set in selecting the primary key in case of more than one candidate key

Constraints on relationship types

- Participant constraint.
  - Total participation: Every entity participates in **at least** one relationship with another entity.



Musicians Example

Exercise 2.5 Notown Records has decided to store information about musicians who perform on its albums (as well as other company data) in a database. The company has wisely chosen to hire you as a database designer (at your usual consulting fee of $2500/day).

- Each musician that records at Notown has an SSN, a name, an address, and a phone number. Poorly paid musicians often share the same address, and no address has more than one phone.
- Each instrument used in songs recorded at Notown has a name (e.g., guitar, synthesizer, flute) and a musical key (e.g., C, B-flat, E-flat).
- Each album recorded on the Notown label has a title, a copyright date, a format (e.g., CD or MC), and an album identifier.
- Each song recorded at Notown has a title and an author.
- Each musician may play several instruments, and a given instrument may be played by several musicians.
- Each album has a number of songs on it, but no song may appear on more than one album.
- Each song is performed by one or more musicians, and a musician may perform a number of songs.
- Each album has exactly one musician who acts as its producer. A musician may produce several albums, of course.

University Example

- Professors have an SSN, a name, an age, a rank, and a research specialty.
- Projects have a project number, a sponsor name (e.g., NSF), a starting date, an ending date, and a budget.
- Graduate students have an SSN, a name, an age, and a degree program (e.g., M.S. or Ph.D.).
- Each project is managed by one professor (known as the project's principal investigator).
- Each project is worked on by one or more professors (known as the project's co-investigators).
- Professors can manage and/or work on multiple projects.
- Each project is worked on by one or more graduate students (known as the project's research assistants).
- Graduate students can work on multiple projects, in which case they will have a (potentially different) supervisor for each one.
- Departments have a department number, a department name, and a main office.
- Departments have a professor (known as the chairman) who runs the department.
- Professors work in one or more departments, and for each department that they work in, a time percentage is associated with their job.
- Graduate students have one major department in which they are working on their degree.