CS3402: Lecture 1

Database Systems

(Sem B, 2024-2025)

Teaching Staff's Information

- Lecturers
  - **Dr. CHEUNG Man Hon** (Michael)
    - Email: mhcheung@cityu.edu.hk
    - Office: YEUNG-Y6412
  - **Dr. HUANG Heqing**
    - Email: heqhuang@cityu.edu.hk
    - Office: YEUNG-Y6429

Course Overview

- Course Format:
  - Face-to-face (F2F) lectures
    - One-hour practice questions
    - Two-hour lecture
  - Tutorial classes **(Start in Week 5 or 6** for seven weeks**)**
    - One-hour F2F tutorial
    - Pattern: 1 time Q&A + 6 labs
    - Bonus for lab attendance

Assessment

- Coursework *-- 40%* :
  - Mid-term *-- 25%*
    - Date: **March 6 (Thur)**
    - Time: Lecture time
    - Format: Open-book exam
  - Homework assignments (3 times) *-- 15%*
  - Lab attendance (At least 5 of the 6 labs) *-- 3% bonus (such that coursework is capped to 40%)*

- Final examination *-- 60%*
  - *Get 30 out of 100 to pass*

Questions about course content?

- Lecture slides & practice questions:
  - Ask questions in lecture time (most efficient)
    - Priority for F2F session
  - Email instructor
    - May be directed to TAs to shorten response time
- Grading:
  - Will let you know who to contact in Canvas later

Teaching Staff (in alphabetical order)

- Teaching assistants (Teaching tutorial sessions):

Name	Email
CUI Yue	yuecui8-c@my.cityu.edu.hk
WANG Lei	lwang678-c@my.cityu.edu.hk

Teaching Staff (in alphabetical order)

- Teaching assistants (Grading):

Name	Email
REN Kean	karen2-c@my.cityu.edu.hk
YANG Ni	niyang3-c@my.cityu.edu.hk
YANG Shuo	syang75-c@my.cityu.edu.hk
YU Jixiang	jixiangyu2-c@my.cityu.edu.hk

Assessment

- **Plagiarism will not be tolerated**
  - But you are welcome to discuss with other students. Just don't copy from one another.
  - https://www.cityu.edu.hk/pvdp/ah/uni-ah-req.htm

Attendance

- If you need to apply for a leave for mid-term and final exams, follow the procedures as laid out in ARRO's webpage (https://www.cityu.edu.hk/arro/asmt/mitg_main.htm) to submit the mitigation request via AIMS no later than 5 working days of the scheduled examination, with supporting documents (e.g., medical certificate).
- If you are ill and cannot attend a lab, obtain a medical note and email it to the instructor of your tutorial section as soon as possible and no later than a week after your absence.

Mark Appeal

- Students are responsible to keep track on their marks.
- Any objection to the course marks should be made to the TA by email within one week of mark announcement. No change will be made afterwards.

Course Materials

- Textbook
  - "**Fundamentals of Database Systems**", **7**th edition, by R. Elmasri and S.B. Navathe, Addison-Wesley.

- Notations may vary in different books. Please stick to the ones used in this lecture notes!

Database is everywhere

- You cannot avoid it and it's everywhere!
- You can say it actually makes the current society and your life work!
- Banking/Credit card /Social Security Info…
- Online shopping/booking…

HANG SENG BANK LTD [HK] https://www.hangseng.com/en-hk/home/

繁 | 简 | Mainland China | Branches & ATMs | SEARCH

恒生銀行 HANG SENG BANK

LOGON to Hang Seng e-Banking

e-Services | Personal | Business | Investment | Insurance and MPF | Cross-Border Banking | Rates | About Us | Links

FPS

Register FPS Now ›

To receive money with JUST a mobile number / email address / FPS ID

Follow global market trends with our investment outlook video

All in "Investment Corner" ›

Investment involves risks.

- Exchange Rates – T/T
- Exchange Rates – Notes
- Gold Prices
- Interest Rates
- HKD Prime Rate
- Property Valuation

Personalized settings

LOGON

- Personal e-Banking
- Business e-Banking
- Hang Seng HSBCnet
- Stock Trading

Personal

Business

About Us

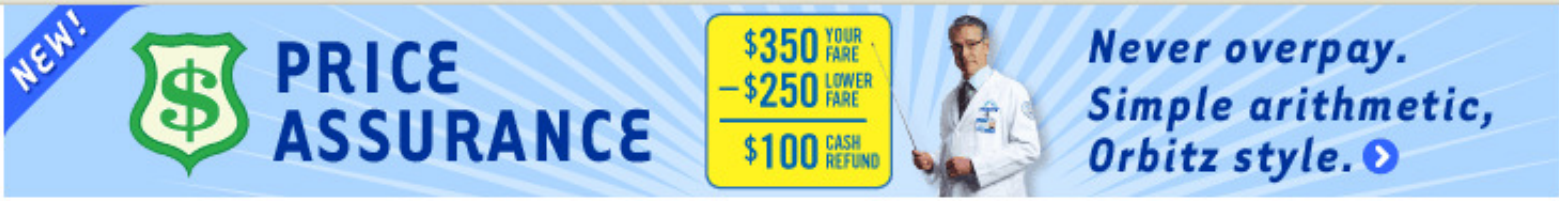
Important Notice

Cheap Travel, Flights, Hotels, Vacations, Car Rentals, Cruise, Activities on Orbitz - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.orbitz.com/

Google Search Bookmarks PageRank Check AutoLink AutoFill Send to



ORBITZ

Welcome to Orbitz .
Sign in | Register now
Site feedback

Search (Atlantis, beach...)
Google™ Custom Search Go

MY STUFF
My Trips | My Account

ORBITZ TLC
Home | Traveler Update | Customer Support

Quick Search | Vacation Packages | Hotels | Flights | Cars & Rail | Cruises | Activities | Deals

- Flight
- Hotel
- Car
- Activities
- Cruises

- Flight + Hotel
- Flight + Car
- Hotel + Car
- Flight + Hotel + Car

Winter Escape
SAVE UP TO 50% ON HOTELS
Get Details

From City name or airport

To City name or airport

Leave
mm/dd/yy
Anytime

Return
mm/dd/yy
Anytime

Travelers (Children or seniors?)
Adult
(18-64)
1

Flight preference
I prefer non-stop flights

Find Flights
Find Flight + Hotel

Expand search options (Multi-city, preferred airlines, etc.)
One-way | Flexible dates

1 2 3 4 5

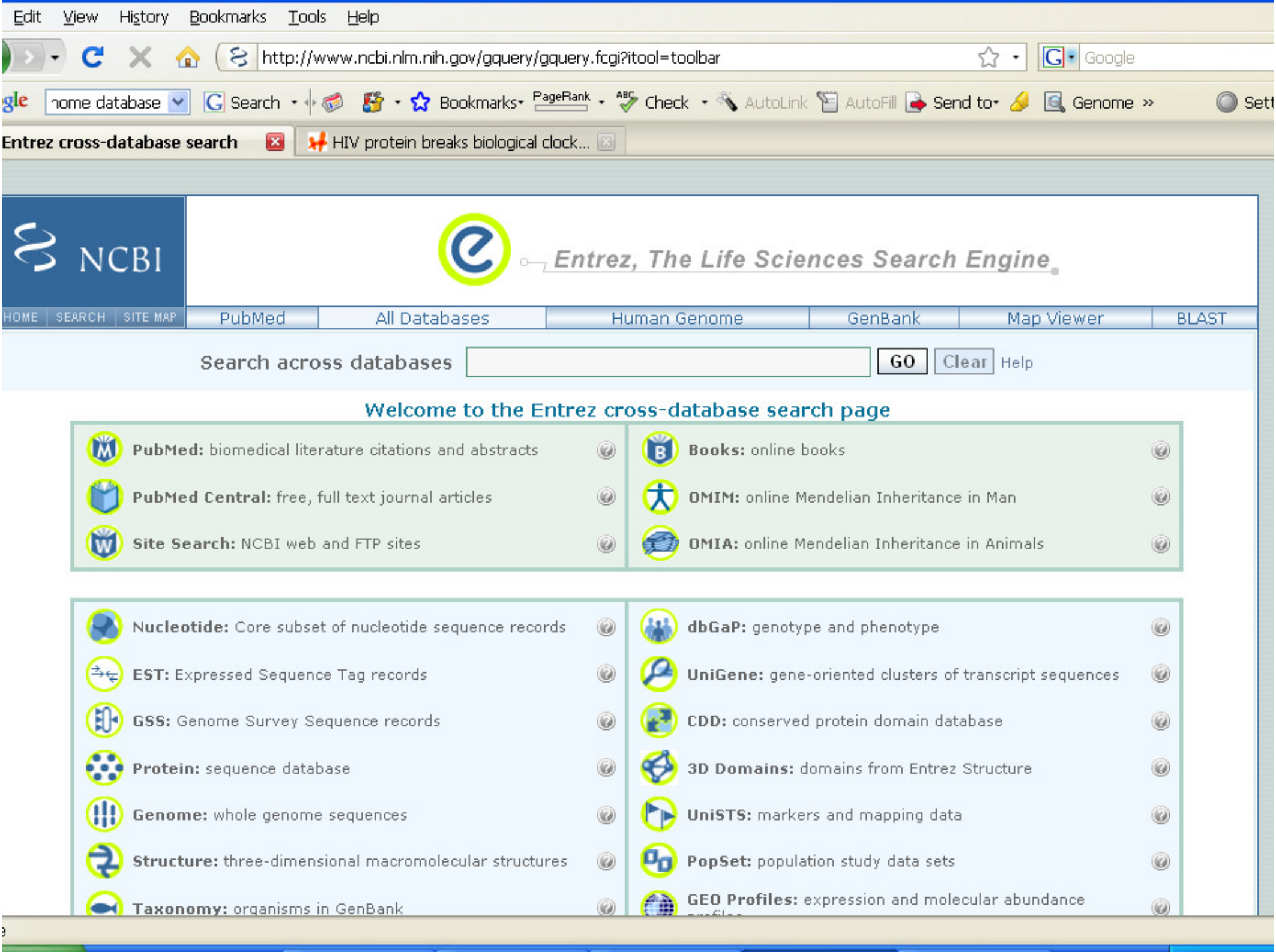
DESTINATIONS
Cancun
Las Vegas
New York City
San Francisco
SEE MORE

INTERESTS
Beach
Romance
Last Minute
Family
SEE MORE

More featured deals View all deals From*

Student? You can travel cheaper! Save
NCL - UP to $500 Free Onboard Spending $199

Done


Edit View History Bookmarks Tools Help
http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi?itool=toolbar
Google
Search Bookmarks PageRank Check AutoLink AutoFill Send to Genome
Entrez cross-database search
HIV protein breaks biological clock...
NCBI
Entrez, The Life Sciences Search Engine
HOME SEARCH SITE MAP
PubMed All Databases Human Genome GenBank Map Viewer BLAST
Search across databases
GO Clear Help
Welcome to the Entrez cross-database search page
PubMed: biomedical literature citations and abstracts
PubMed Central: free, full text journal articles
Site Search: NCBI web and FTP sites
Books: online books
OMIM: online Mendelian Inheritance in Man
OMIA: online Mendelian Inheritance in Animals
Nucleotide: Core subset of nucleotide sequence records
EST: Expressed Sequence Tag records
GSS: Genome Survey Sequence records
Protein: sequence database
Genome: whole genome sequences
Structure: three-dimensional macromolecular structures
Taxonomy: organisms in GenBank
dbGaP: genotype and phenotype
UniGene: gene-oriented clusters of transcript sequences
CDD: conserved protein domain database
3D Domains: domains from Entrez Structure
UniSTS: markers and mapping data
PopSet: population study data sets
GEO Profiles: expression and molecular abundance

Amazon.com: Online Shopping for Electronics, Apparel, Computers, Books, DVDs & more - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.amazon.com/

Google Search Bookmarks PageRank Check AutoLink AutoFill Send to

amazon.com

Hello, Ruoming Jin. We have recommendations for you. (Not Ruoming?)

FREE 2-Day Shipping, No Minimum Purchase

Ruoming's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards | Your Account | Help

Shop All Departments

- Books
- Movies, Music & Games
- Digital Downloads
- Computers & Office
- Electronics
- Home & Garden
- Grocery, Health & Beauty
- Toys, Kids & Baby
- Apparel, Shoes & Jewelry
- Sports & Outdoors
- Tools, Auto & Industrial

Search All Departments GO Cart Your Lists

100% More

Books, Newspapers, and Blogs

We've doubled our selection since launch to over 200,000 titles—all available wirelessly in under 60 seconds:

- *New York Times* bestsellers are $9.99, unless otherwise marked
- Expanded newspaper selection, including *The New York Times*, *Wall Street Journal*, and *Financial Times*
- More magazines and blogs, from *Newsweek* to *Time* to *Gawker*

amazonkindle Learn more

Amazon Daily BLOG

5 posts since yesterday

Posts for Ruoming

Treat Yourself

Combinatorial Optimization

From the reviews: What a book! Actually it is three books which come... Read more

~~$135.00~~ $107.51

See more in your Wish List

Get Your Motor Running

Explore Our New Motorcycle & ATV Store

Shop now at Amazon.com

Check This Out

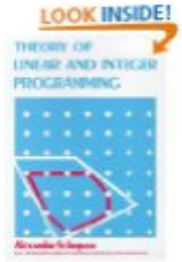
Tax Central
Find tax software, books, and more

Cast Your Vote
See our Top 10 New Year's Resolutions. Which one's for you?

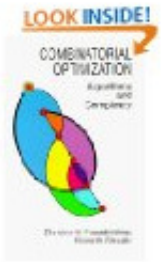
Holiday Returns
Visit our self-service Online Returns Center.

For the Photography Fan
Zoom in on top

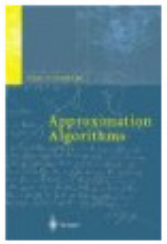
Frequently Bought with Items in Your Wish List



Theory of Linear and Integer Programming
Paperback by Alexander



Combinatorial Optimization...
Paperback by Christos H.



Approximation Algorithms Hardcover by Vijay V. Vazirani

Done

Course Objectives

- ER model: characterize relationships among entities
- Relational model: transform from ER diagram to tables
- Normal Forms: how to design good tables
- SQL: language for writing queries
- Relational Algebra: logical way to represent queries
- File Organization: provide file level structure to speed up query
- Transactions and Concurrency Control: handle concurrent operations and guarantee correctness of the database
- Query Optimization: transform queries into more efficient ones

Introduction to DB Systems

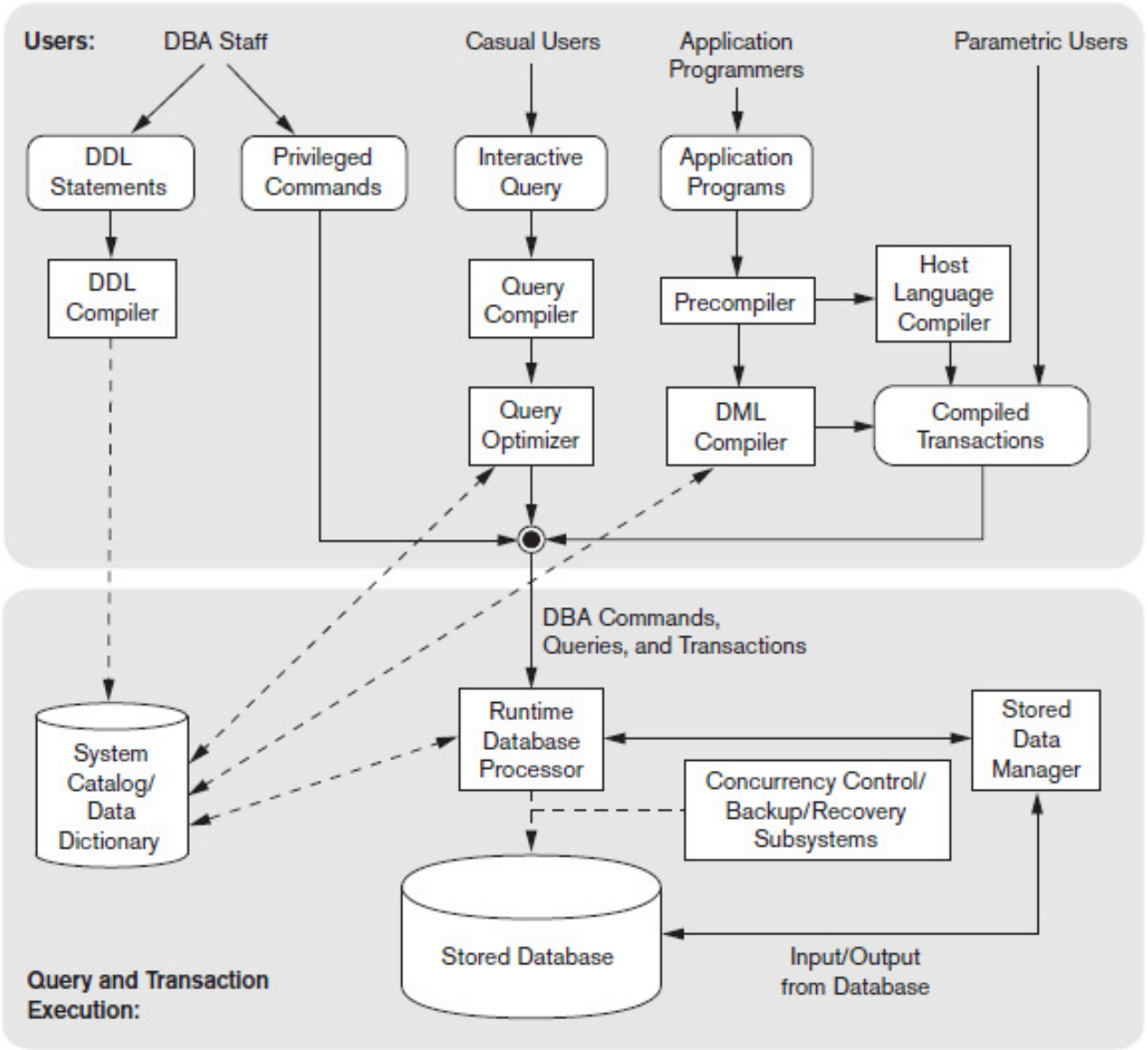
- What is a Database (DB)?
  - A non-redundant, persistent collection of logically-related records/files that are structured to support various processing and retrieval needs.

- Database Management System (DBMS)
  - A set of software programs for creating, storing, updating, and accessing the data of a DB.
  - E.g.: Oracle, Mysql，Oceanbase

DB System Architecture

- *Data Query Language (DQL)*
- a language used to make queries in databases
- e.g. search records with giving conditions (sex="Female")
- *Data Manipulation Language (DML)*
- a language that enables users to manipulate data
- e.g. insert or delete records
- *Data Definition Language (DDL)*
- a language for defining DB schema
- e.g., create, modify, and remove database objects such as tables, indexes, and users.

DB System Architecture



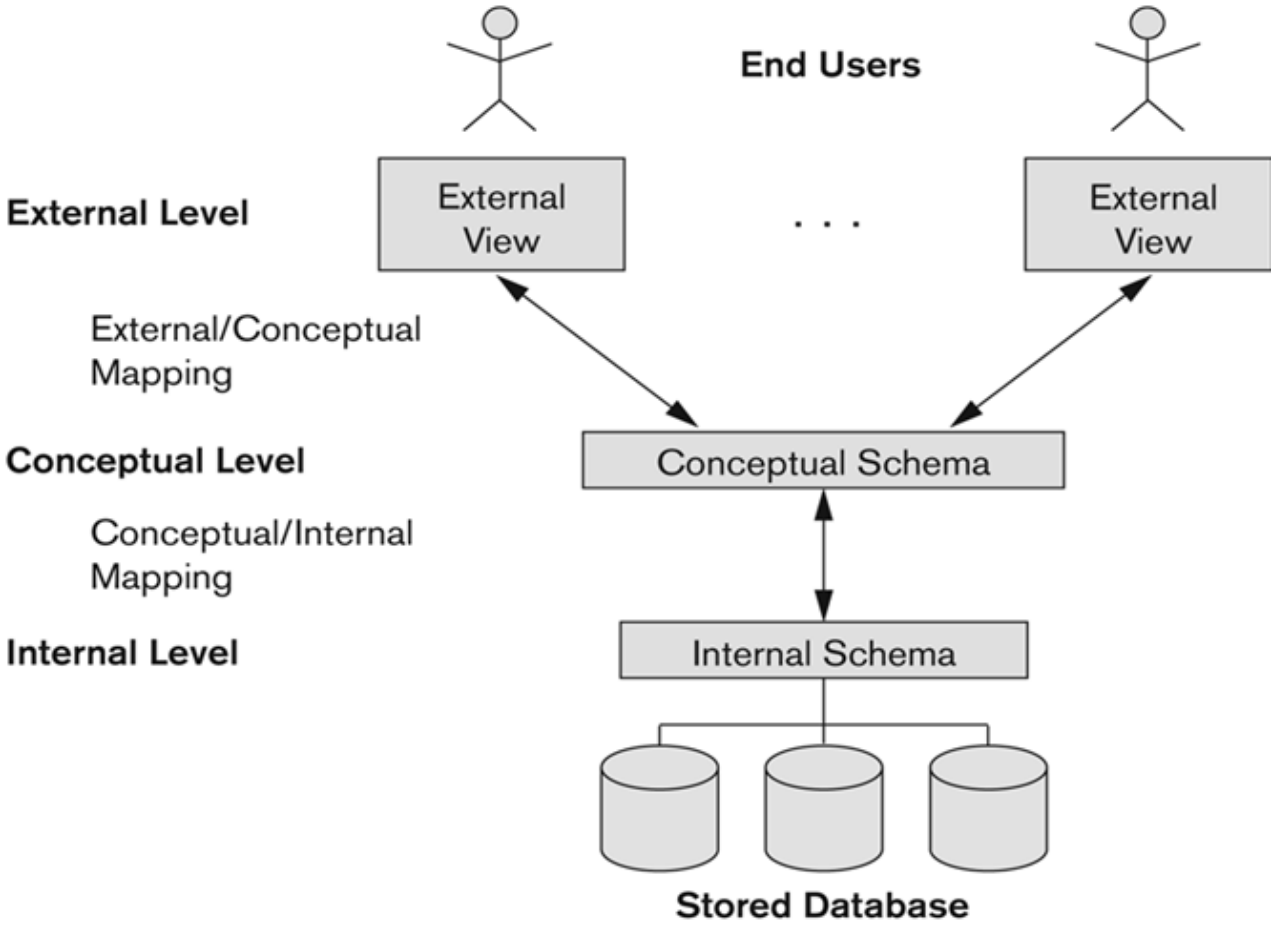
DB System Architecture

- *Database Administrator (DBA)*
  - DBA is the person who has central control over the DB
  - Main functions of DBA:
    - schema definition
    - storage structure and access method definition
    - schema and physical organization modification
    - granting of authorization for data access
    - integrity constraint specification

DB System Architecture

- *Database Users*
  - Application Programmers:
    - Writing embedded DML in a host language
  - Interactive Users (i.e., Causal Users):
    - Using query languages
  - Naive Users (i.e., Parametric Users):
    - Running application programs

Three-schema architecture



- Goal: Separate the user applications from the physical database
- Schema can be defined in 3 levels

Three-schema architecture

- Physical/internal level: **internal schema** uses a physical data model and describes the complete details of data storage and access paths for the database.

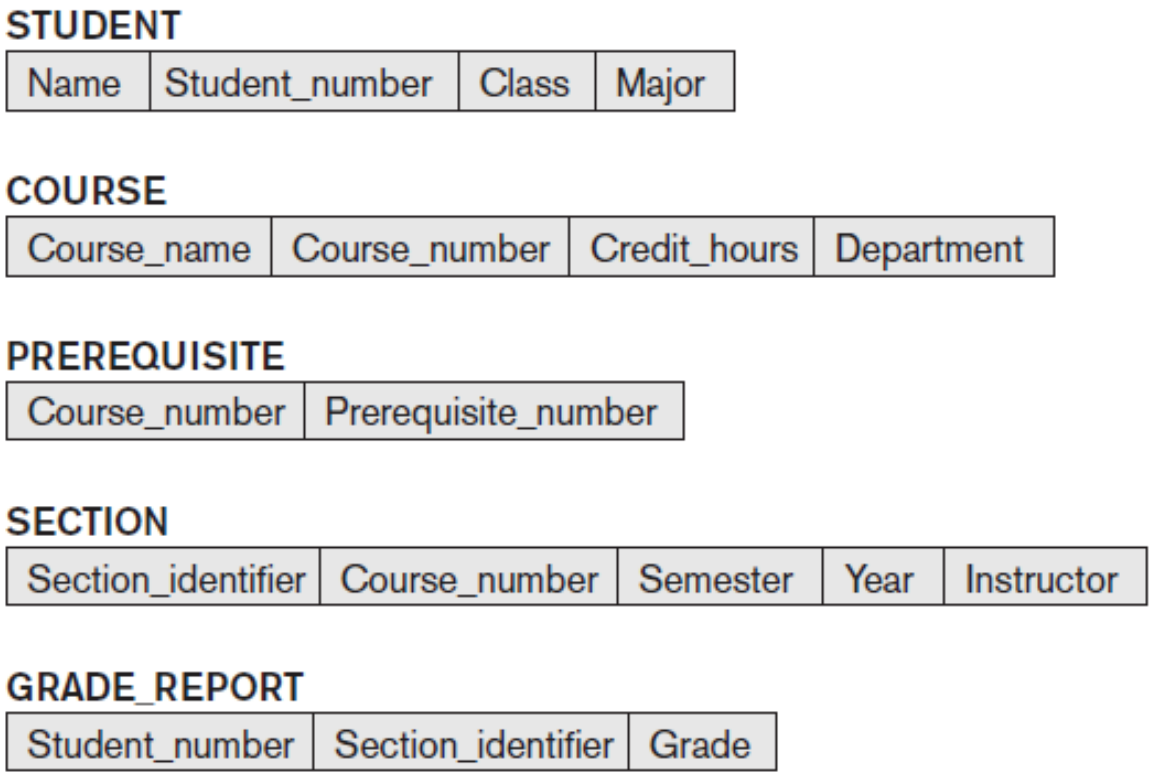
Data Item Name	Starting Position in Record	Length in Characters (bytes)
Name	1	30
Student_number	31	4
Class	35	1
Major	36	4

Figure 1.4
Internal storage format for a STUDENT record, based on the database catalog in Figure 1.3.

Three-schema architecture

- Conceptual level: **conceptual schema** describes the structure of the whole database for a community of users and hides the details of physical storage structures.

Figure 2.1
Schema diagram for the database in Figure 1.2.



Three-schema architecture

- View/external level: **external schema** describes the part of the database that a particular user group is interested in and hides the rest from that group.

TRANSCRIPT

Student_name	Student_transcript				
	Course_number	Grade	Semester	Year	Section_id
Smith	CS1310	C	Fall	08	119
	MATH2410	B	Fall	08	112
Brown	MATH2410	A	Fall	07	85
	CS1310	A	Fall	07	92
	CS3320	B	Spring	08	102
	CS3380	A	Fall	08	135

(a)

COURSE_PREREQUISITES

Course_name	Course_number	Prerequisites
Database	CS3380	CS3320
		MATH2410
Data Structures	CS3320	CS1310

(b)

Figure 1.5
Two views derived from the database in Figure 1.2. (a) The TRANSCRIPT view. (b) The COURSE_PREREQUISITES view.

Data Independence

- Data Independence: the ability to modify a schema definition in one level without affecting a schema in the next higher level
- two types of data independence:
  - ***logical data independence***

    -- the ability to modify the conceptual schema without causing the application programs to be rewritten
  - ***physical data independence***

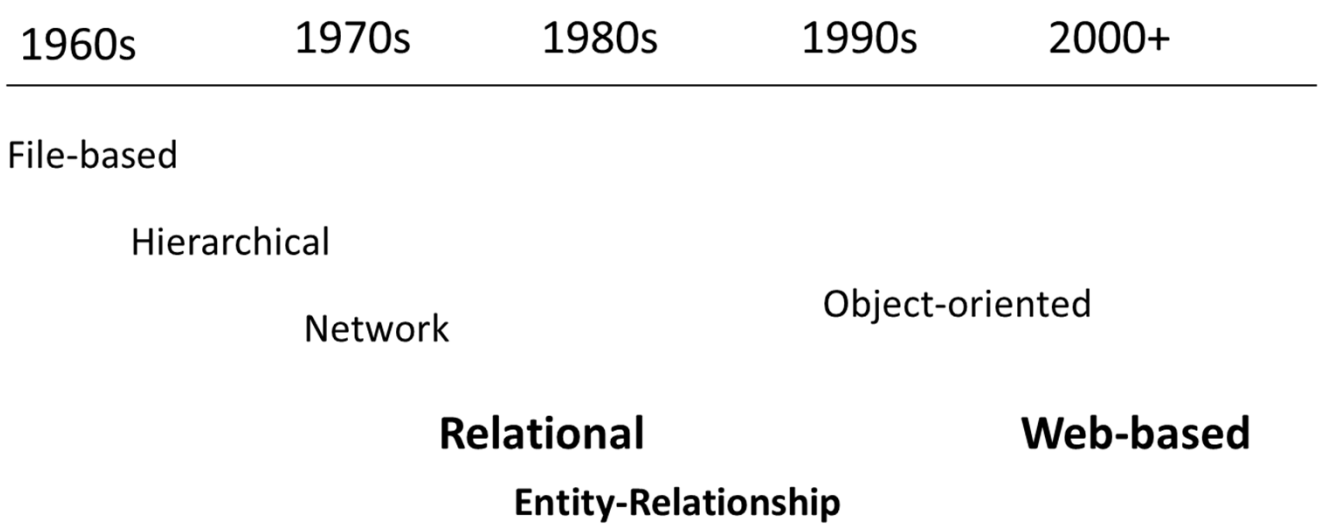
    -- the ability to modify the physical schema without altering the conceptual schema and thus, without causing the application programs to be rewritten

Data Models

- Data Model (conceptual level)
  - A collection of conceptual tools for describing data, data relationships, operations, and consistency constraints
  - the "core" of a database

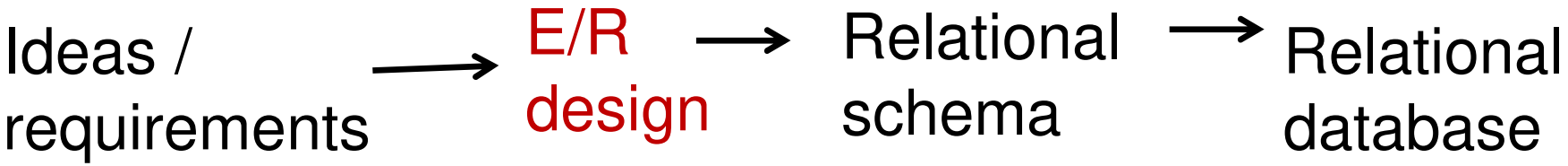
Evolution of Data Models

- Timeline



The Entity-Relationship Model

- Preliminaries
  - Proposed by P. Chen in 1976
  - Direct, easy-to-understand graphical notation
  - Translates readily to relational schema for database design



- **Three basic concepts:**

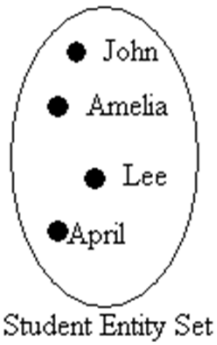
  Entity, Attribute, Relationship

ER Model Concepts

- **Entity**
  - a distinguishable object with an independent existence

    Example: John Chan, CityU, HSBC, …
- **Entity Set**
  - a set of entities of the same type

    Example: Student, University, Bank, ...



This student entity set has four entities in it: John, Amelia, Lee, and April

Figure 1-6. Entity set and its entities

ER Model Concepts

- **Attribute**(Property) -- a piece of information describing an entity
  - *Example*: Name, ID, Address, Sex are attributes of a student entity
  - Each attribute can take a value from a domain

    Example: Name $\in$ Character String,

    ID $\in$ Integer, ...
  - Formally, an attribute **A** is a function which maps from an entity set **E** into a domain **D**:

    $$\mathbf{A: E \rightarrow D}$$

Types of Attributes

- Simple attributes
  - Attributes that are not divisible.
  - Example: SSN, gender
- Composite attributes
  - The attribute may be composed of several components.
  - Composition may form a hierarchy where some components are themselves composite

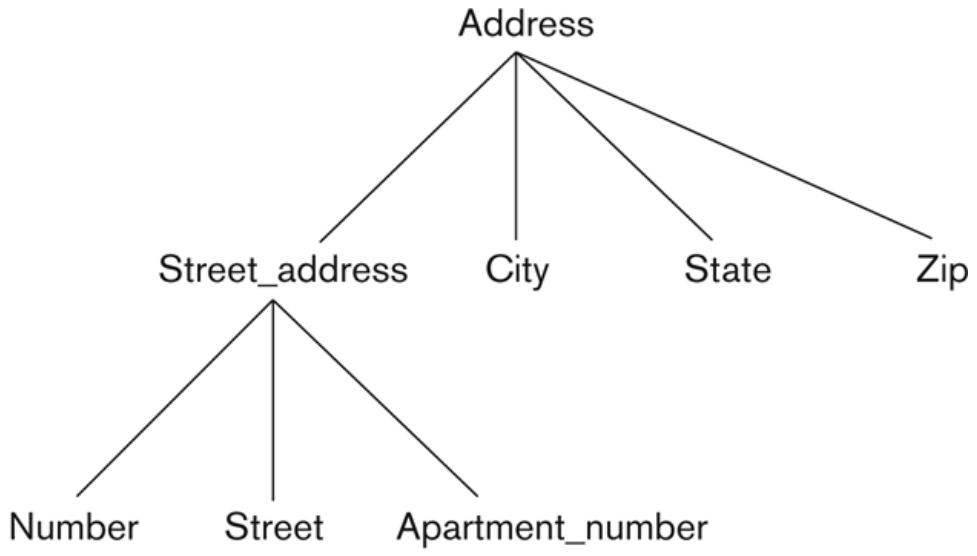


Figure 3.4
A hierarchy of composite attributes.

Types of Attributes

- Single-valued attributes
  - Most attributes have a single value for a particular entity
    - E.g., Age of a person
- Multi-valued attributes
  - An entity may have multiple values for that attribute.
    - E.g., Color of a CAR: {red, black} (i.e., two-tone car)

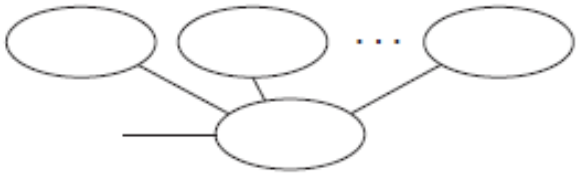
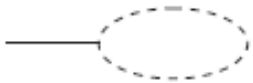
Complex Attributes

- In general, composite and multi-valued attributes may be nested to any number of levels
- For example, PreviousDegrees of a STUDENT is a composite multi-valued attribute denoted by {PreviousDegrees (College, Year, Degree, Field)}:
  - Multi-valued: Multiple PreviousDegrees values can exist
  - Composite: Each has four subcomponent attributes (College, Year, Degree, Field)

Key Attributes

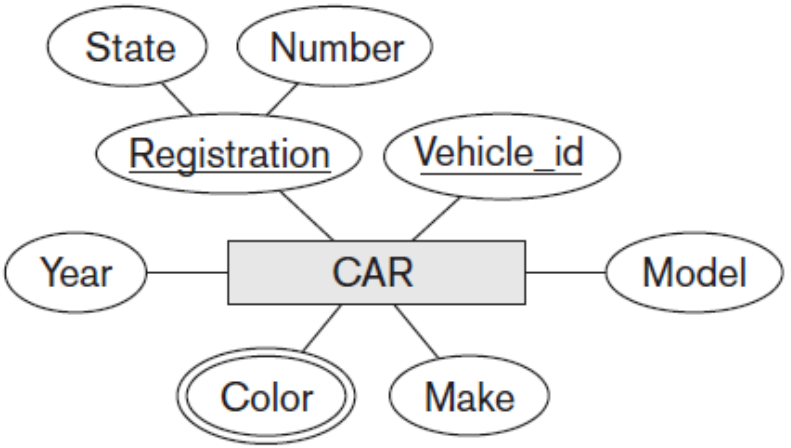
- An important constraint on the entities of an entity type is the key constraint on attributes.
- Key attribute: One or more attributes whose values are distinct for each individual entity in the entity set.
  - Its values can be used to identify each entity uniquely.
- Example: Consider a STUDENT entity type
  - Is student name a key attribute?
  - Is student ID a key attribute?

ER Model Diagram

Symbol	Meaning
	Entity
	Attribute
	Key Attribute
	Multivalued Attribute
	Composite Attribute
	Derived Attribute

Key Attributes (cont'd)

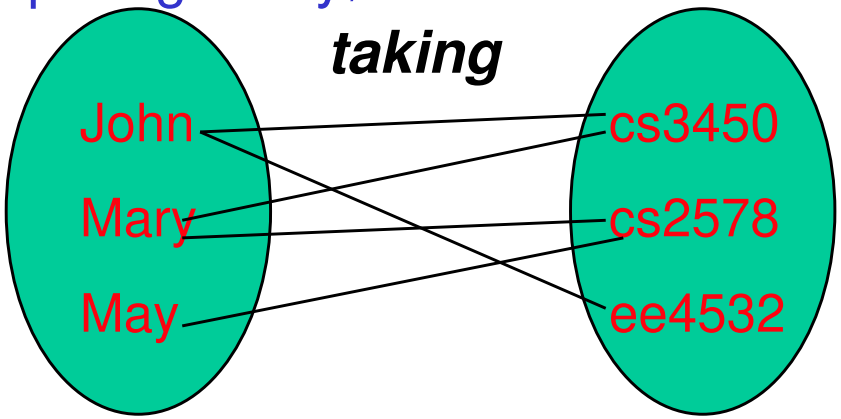
- Sometimes several attributes together form a key
  - Define a composite attribute and designate it as a key attribute of the entity type.
- Notice that such a composite key must be minimal:
  - All component attributes must be included in the composite attribute to have the uniqueness property.
  - Superfluous attributes must not be included in a key.
- Some entity types have more than one key attribute.



ER Model Concepts

- **Relationship** -- an association among several entities
  - *Example*: Patrick and Eva are friends
    John is taking cs3450
- a relationship can carry attributes: properties of the relationship
  - Example: John takes cs3450 with a grade of B+

- **Relationship Set** -- a set of relationships of the same type (same attribute, same participating entity, same constraints)
  - Example:



  - Formally, a relationship **R** is a subset of:
    { (e1, e2, …, ek) | e1 ∈ E1, **e2** ∈ E2, …, **ek** ∈ Ek) }

Example COMPANY Database

- We need to create a database schema (definition) based on the following (simplified) application requirements of the COMPANY Database:
  - R1. The company is organized into DEPARTMENTs
  - R2. Each DEPARTMENT has a unique name, unique number and an EMPLOYEE who *manages* the department
  - R3. We keep track of the start date of the department manager. A department may have several locations
  - R4. Each DEPARTMENT controls/has a number of PROJECTs
  - R5. Each project has a unique name, unique number and is located at a single location

Example COMPANY Database

- R6. The database will store each EMPLOYEE's social security number (ssn), name(first name, last name and middle init), address, salary, sex, and birthdate
  - a. Each employee *works for* one department but may *work on* several projects
  - b. The DB will keep track of the number of hours per week that an employee currently works on each project
  - c. It is required to keep track of the *direct supervisor* of each employee

- R7. Each employee may *have* a number of DEPENDENTs
  - a. For each dependent, the DB keeps a record of name, sex, birthdate, and relationship to the employee

Initial Conceptual Design of Entity Sets

- Based on the requirements, we can identify four initial entity sets in the COMPANY database:
  - DEPARTMENT
  - PROJECT
  - EMPLOYEE
  - DEPENDENT

- The initial attributes shown are derived from the requirements description:

Initial Design of Entity Sets:
EMPLOYEE, DEPARTMENT, PROJECT, DEPENDENT

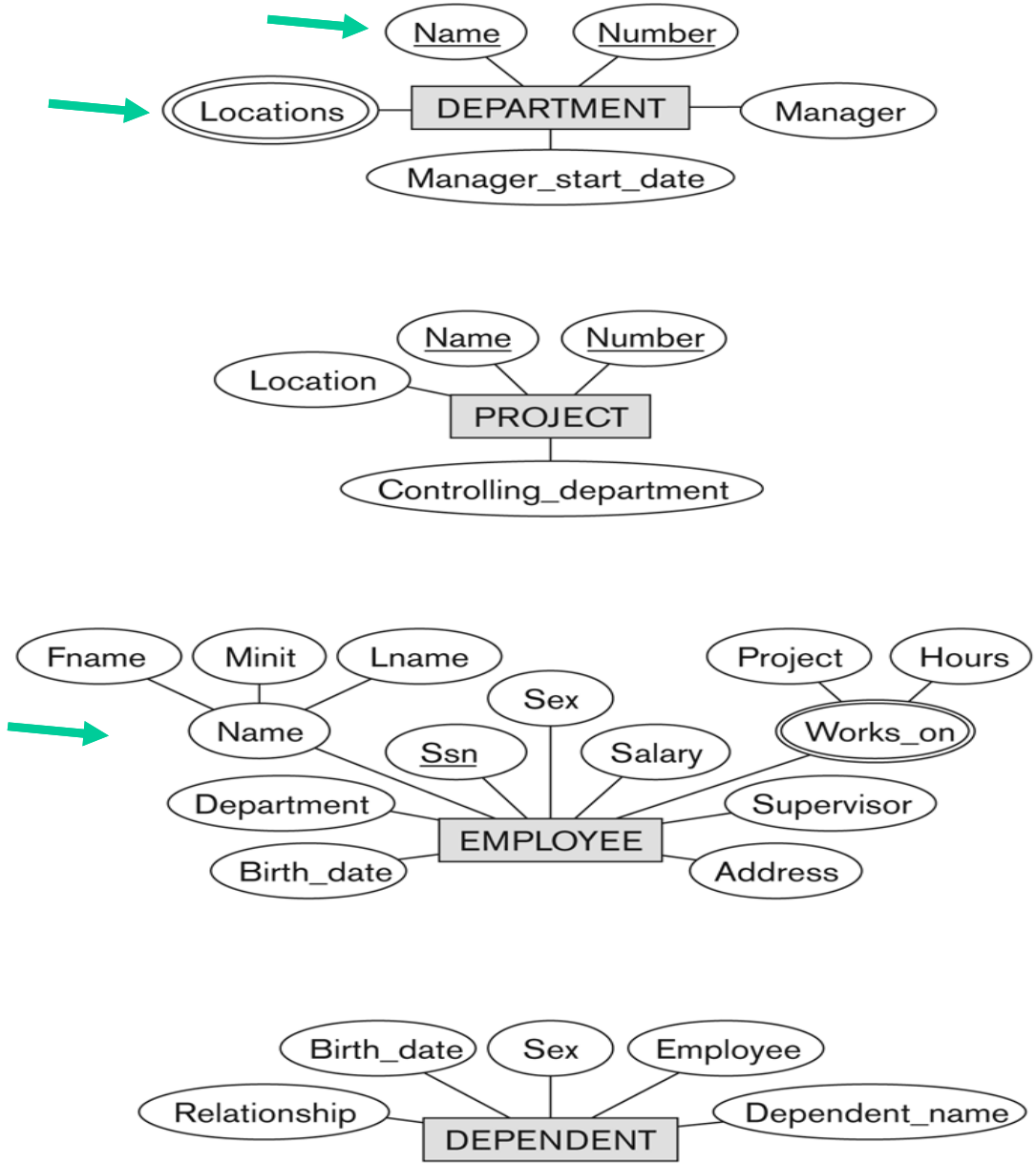


Figure 3.8
Preliminary design of entity types for the COMPANY database. Some of the shown attributes will be refined into relationships.

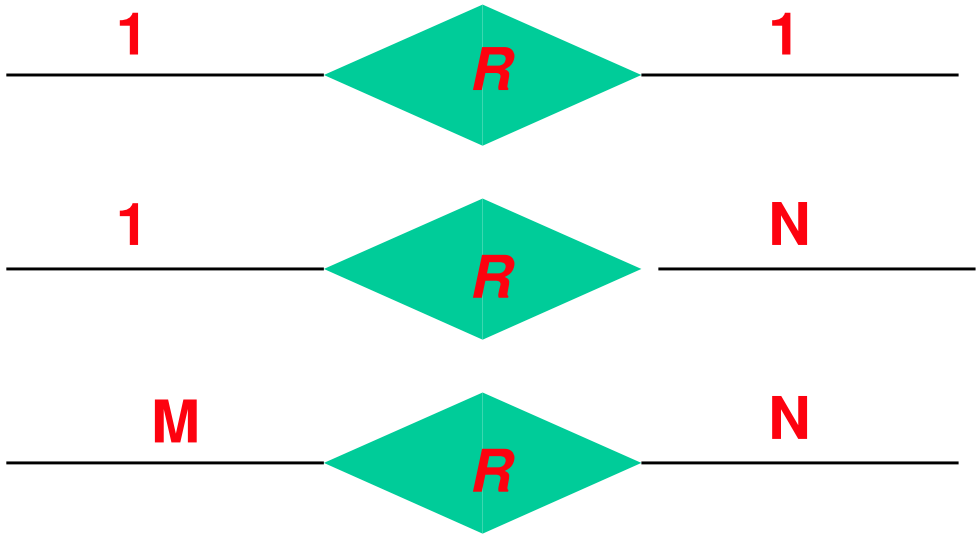
Refining the initial design by introducing relationships

- Find relationships relating two or more distinct entities/entity types with a specific meaning

- By examining the requirements, six relationship types are identified:
  - EMPLOYEE *works on* PROJECT
  - EMPLOYEE *works for* DEPARTMENT
  - EMPLOYEE *manages* DEPARTMENT
  - EMPLOYEE *supervises* EMPLOYEE
  - DEPENDENT *depends on* EMPLOYEE
  - DEPARTMENT *controls* PROJECT

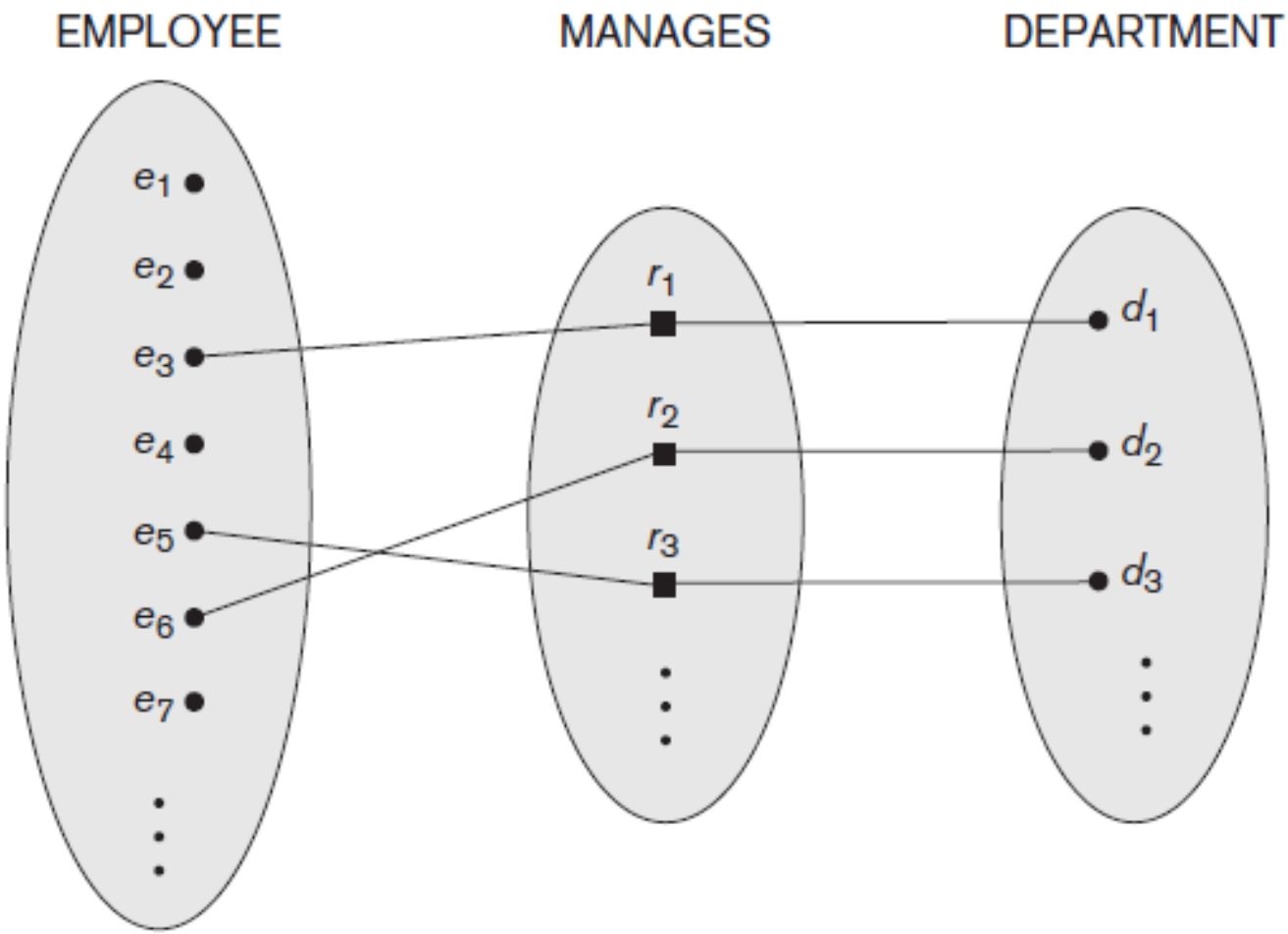
Constraints on relationship

- Cardinality ratio: specifies the maximum number of relationship instances that an entity can participate in
- Possible cardinality ratios for binary relationship types: 1:1, 1:N, N:1, or M:N

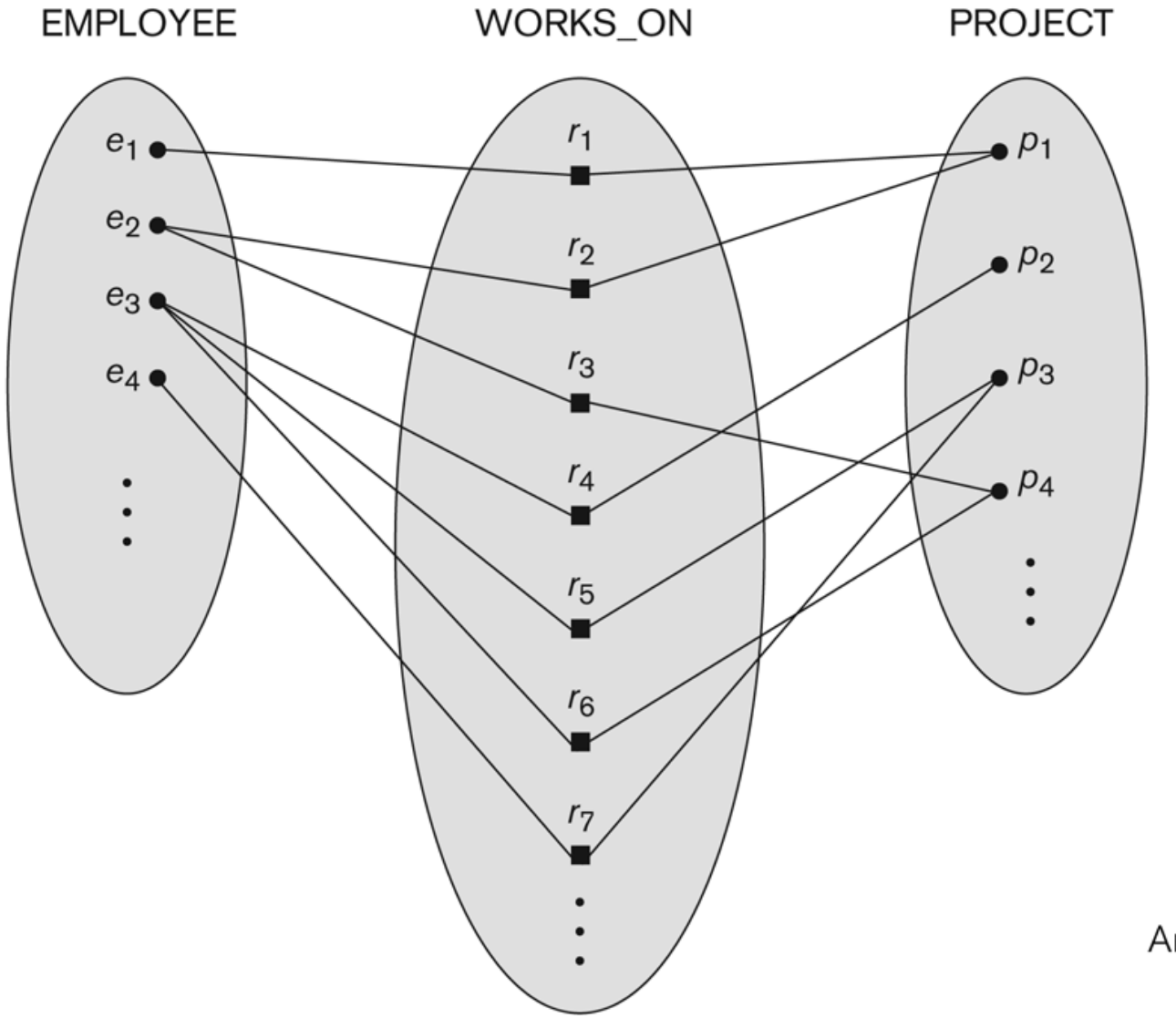


One-to-one (1:1) Relationship

Figure 3.12
A 1:1 relationship, MANAGES.



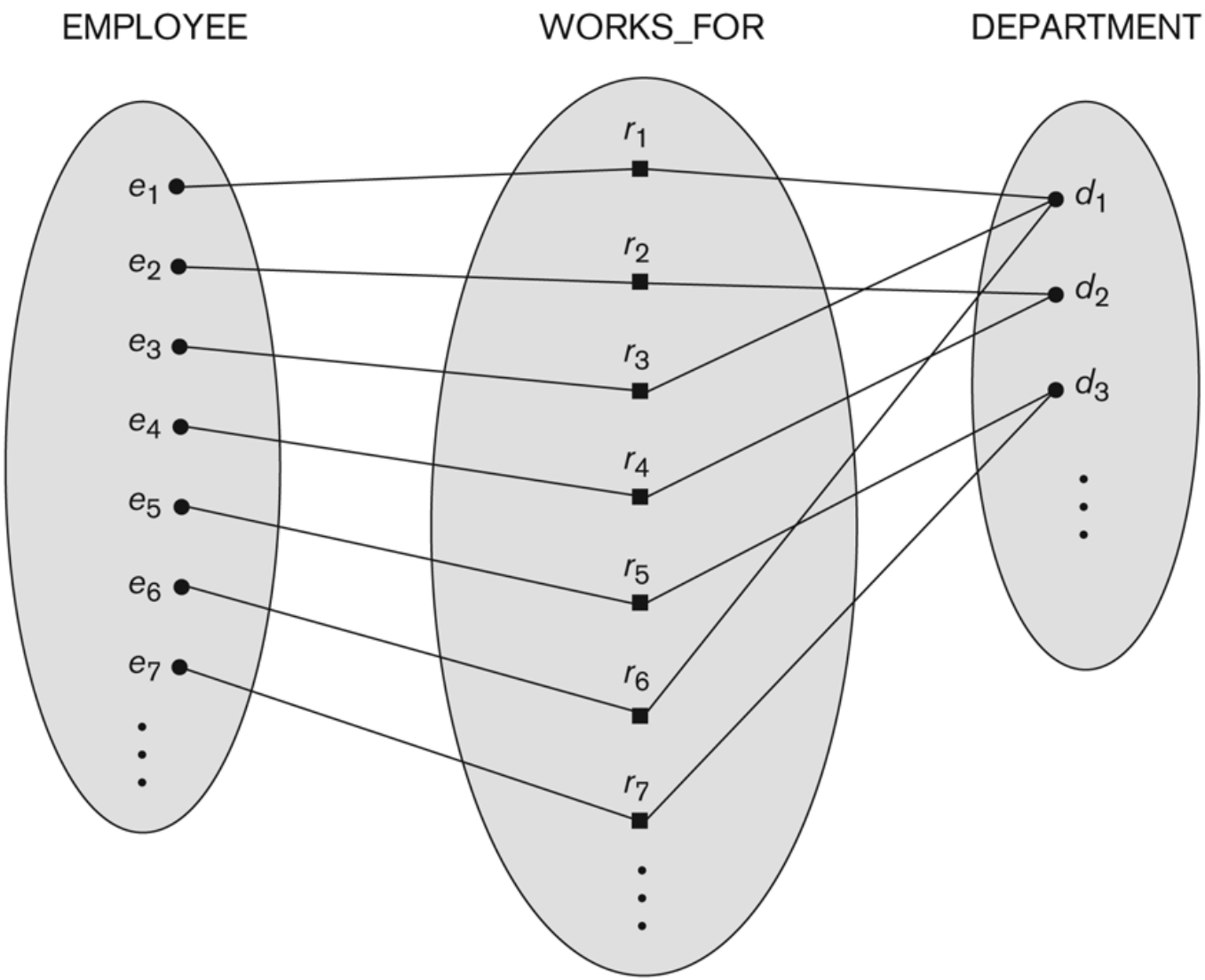
Many-to-many (M:N) Relationship



Many employee work on many projects

Figure 3.13
An M:N relationship, WORKS_ON.

Many-to-one (N:1) Relationship

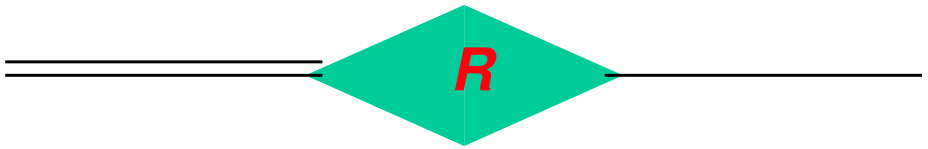


Many employees in the same department

Figure 3.9
Some instances in the WORKS_FOR relationship set, which represents a relationship type WORKS_FOR between EMPLOYEE and DEPARTMENT.

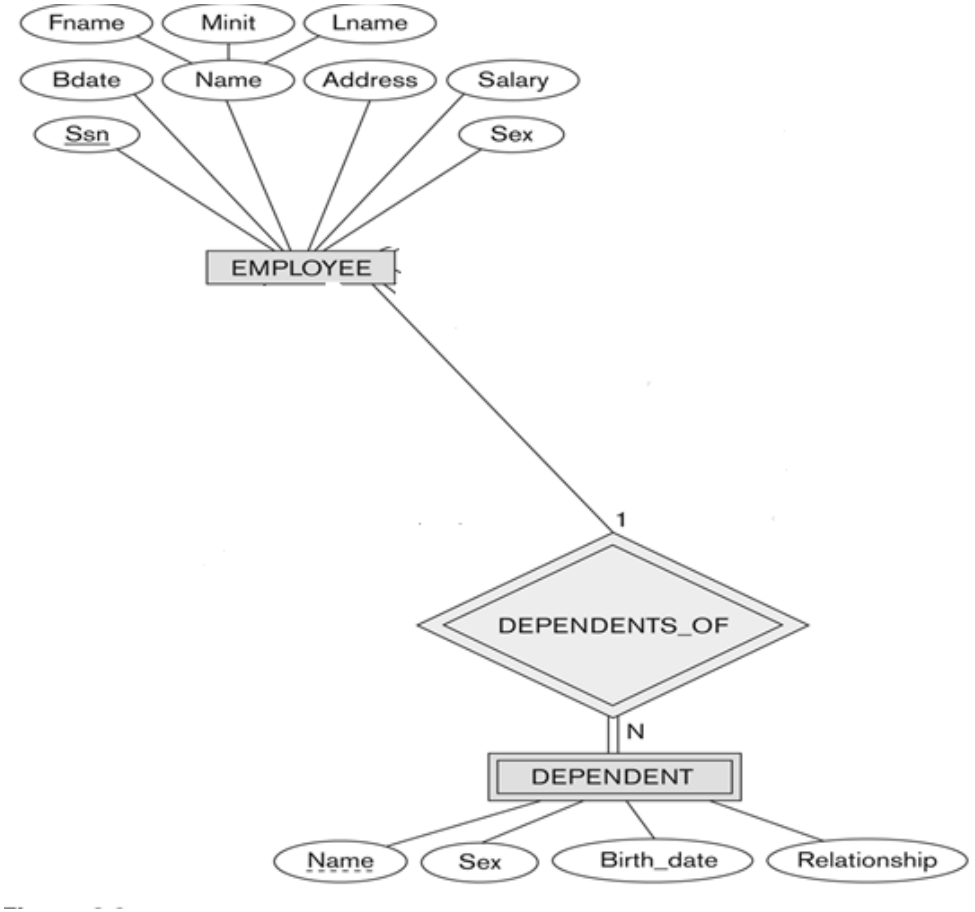
Constraints on relationship

- Participation constraint:
  - specifies the minimum number of relationship instances that an entity can participate in
  - Two types of participation constraints:
    - Total participation: shown by double line
    - Partial participation: shown by single line



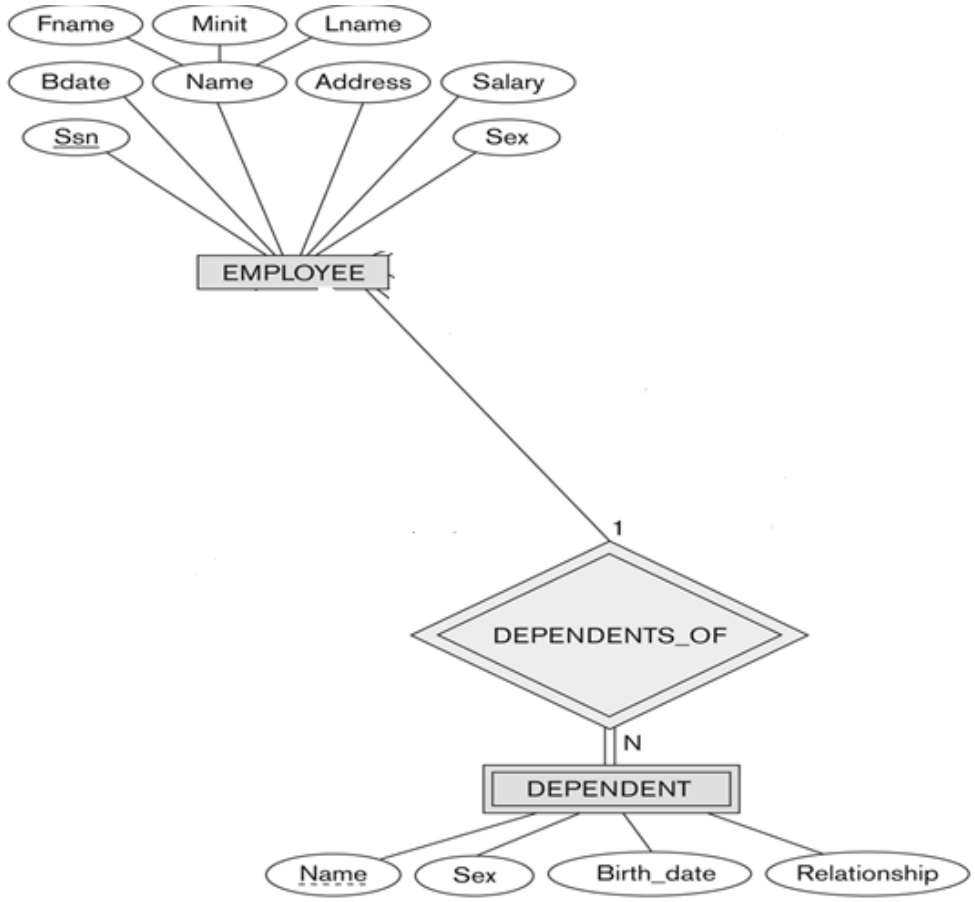
Weak Entity Types

- Strong entity types: Regular entity types that have key attribute(s).
- Weak entity types: Entity types that do **not** have key attributes of their own.
- Entities belonging to a weak entity type are identified by being related to specific entities from an owner entity type in combination with one of their attribute values.
- Identifying relationship of weak entity type: The relationship type that relates a weak entity type to its owner.



Weak Entity Types (cont'd)

- A weak entity type always has a total participation constraint (existence dependency) with respect to its identifying relationship.
  - Reason: a weak entity cannot be identified without an owner entity.
- A weak entity type normally has a partial key, which is the attribute that can uniquely identify weak entities that are related to the same owner entity



ER DIAGRAM – Relationship Types are:

WORKS_FOR, MANAGES, WORKS_ON, CONTROLS, SUPERVISION, DEPENDENTS_OF

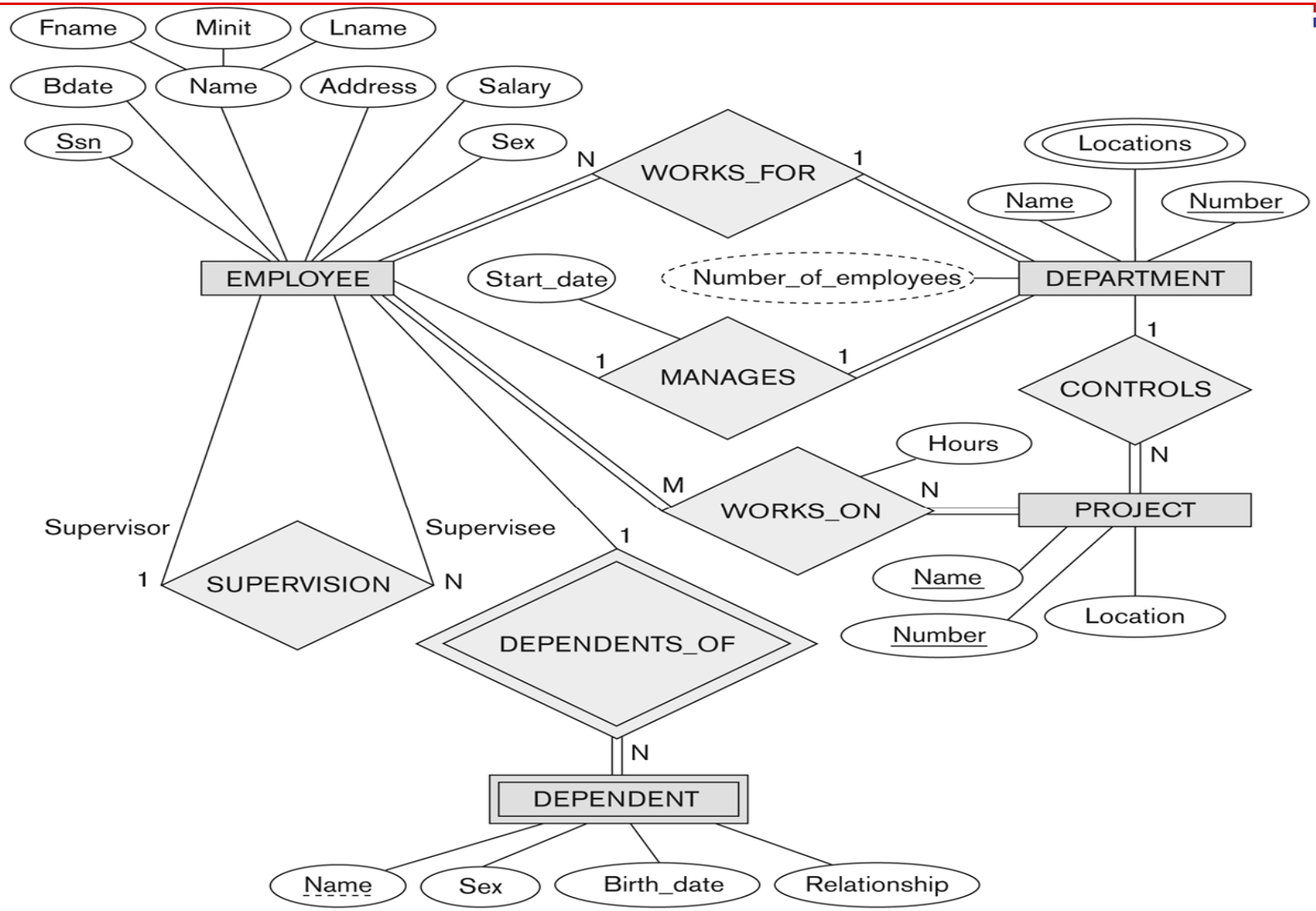
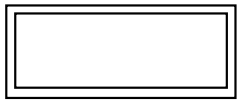
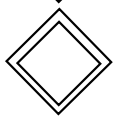
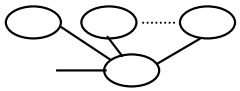
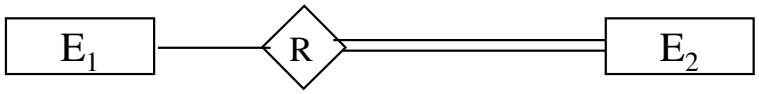
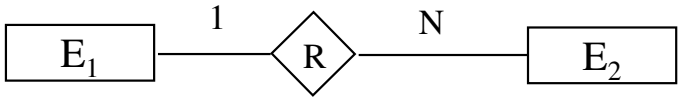
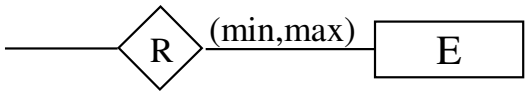


Figure 3.2
An ER schema diagram for the COMPANY database. The diagrammatic notation is introduced gradually throughout this chapter.

Summary of ER-Diagram Notation

Symbol	Meaning
	ENTITY TYPE
	WEAK ENTITY TYPE
	RELATIONSHIP TYPE
	IDENTIFYING RELATIONSHIP TYPE
	ATTRIBUTE
	KEY ATTRIBUTE
	MULTIVALUED ATTRIBUTE
	COMPOSITE ATTRIBUTE
	DERIVED ATTRIBUTE
	TOTAL PARTICIPATION OF E_2 IN R
	CARDINALITY RATIO 1:N FOR E_1:E_2 IN R
	STRUCTURAL CONSTRAINT (min, max) ON PARTICIPATION OF E IN R

Summary

- ER diagram
  - Notations (e.g., entity, attribute, relationship)
  - Key attributes
  - Cardinality ratio
  - Participation constraint

References

- 7e
  - Ch. 2
  - Ch. 3