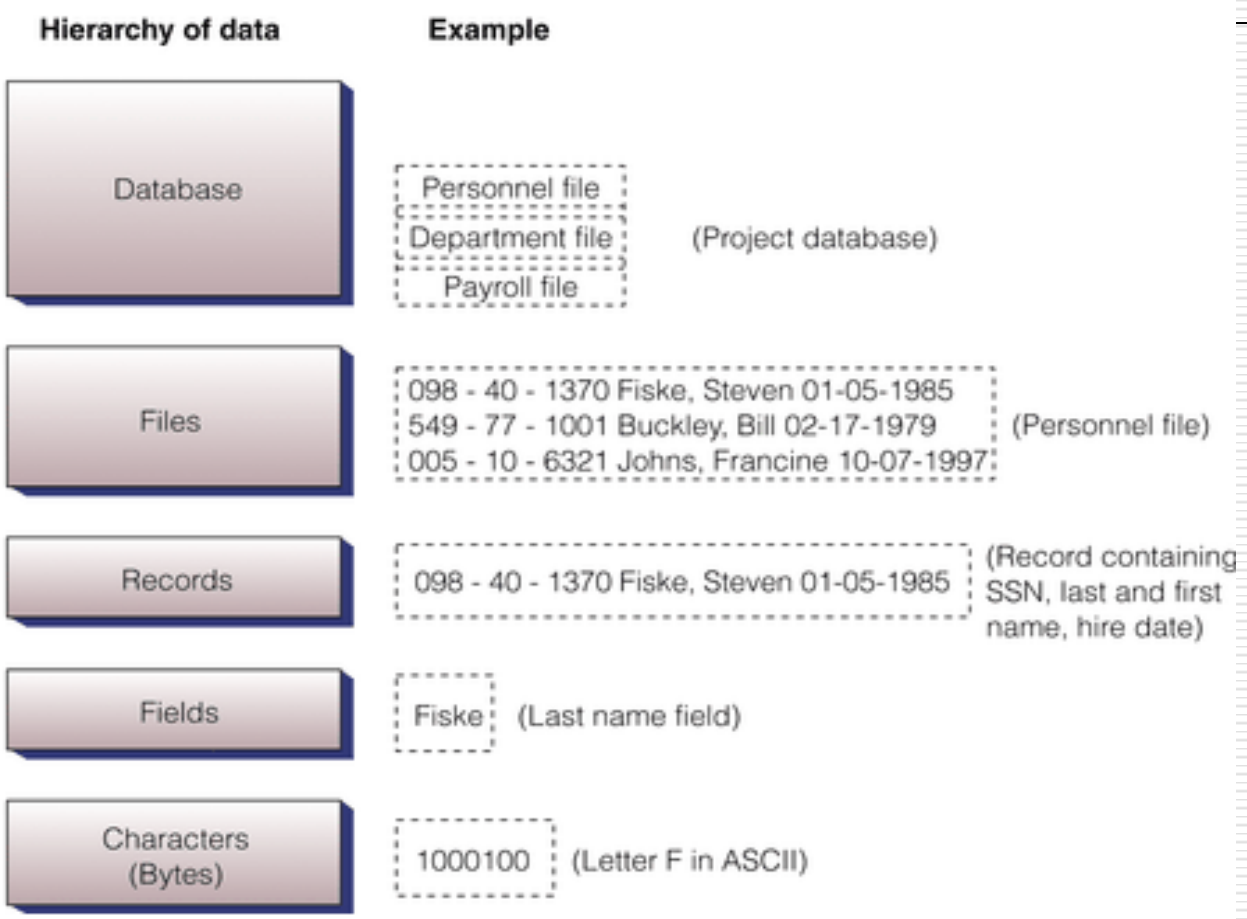
Introduction to Computing

Lectured by: Dr. Pham Tran Vu

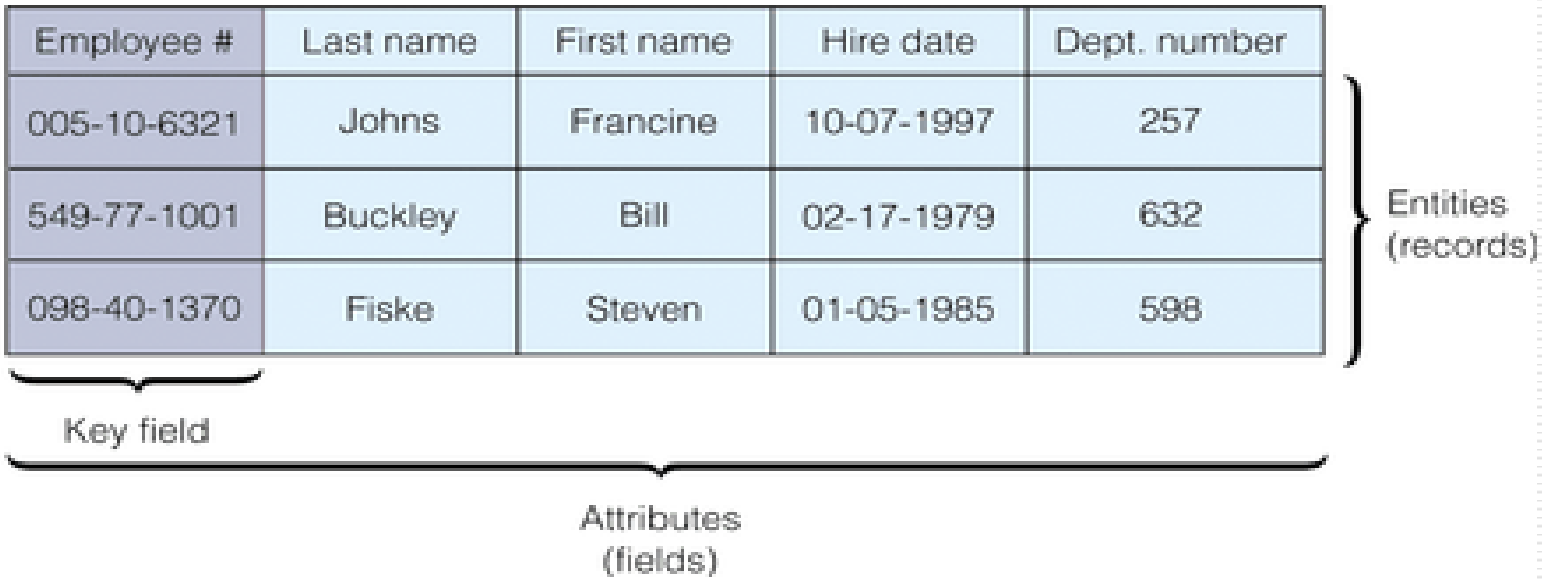
t.v.pham@cse.hcmut.edu.vn

Databases

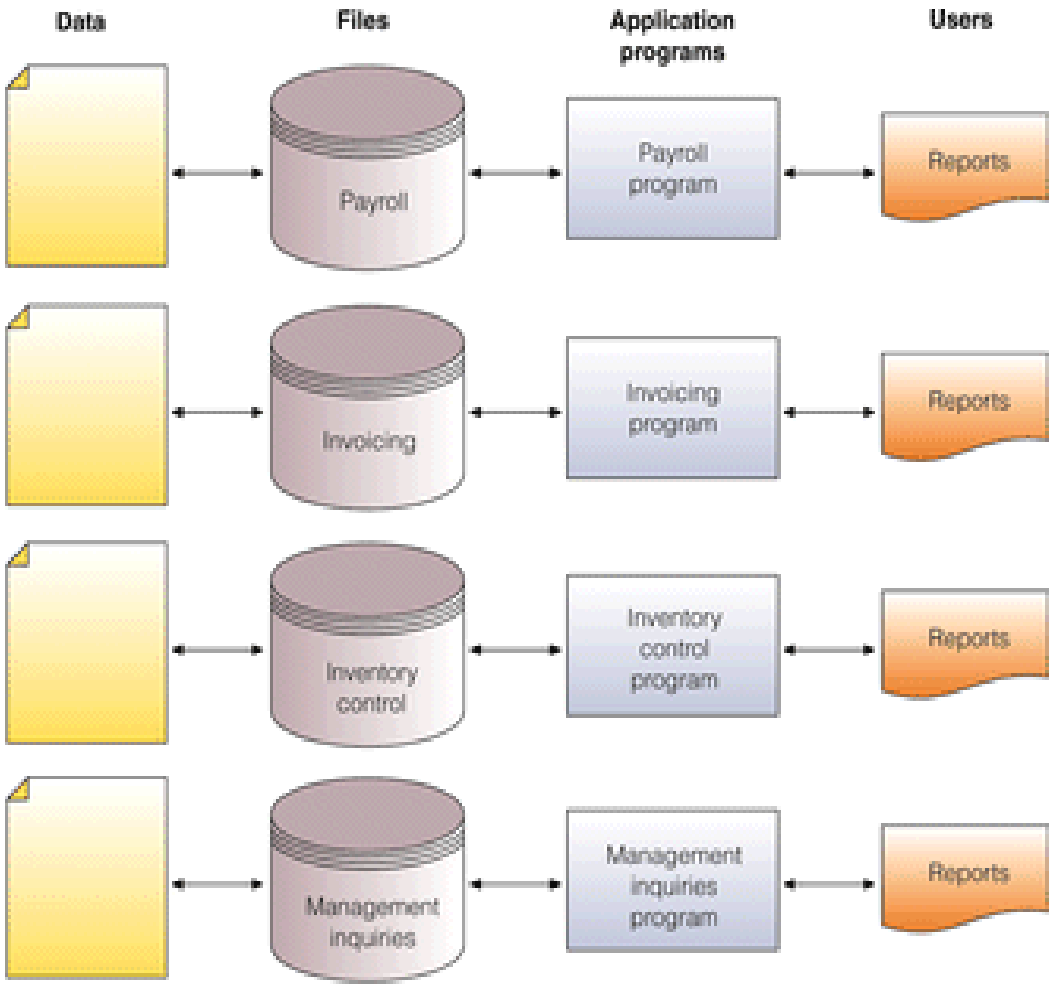
The Hierarchy of Data



Keys and Attributes



The Traditional Approach To Data Management

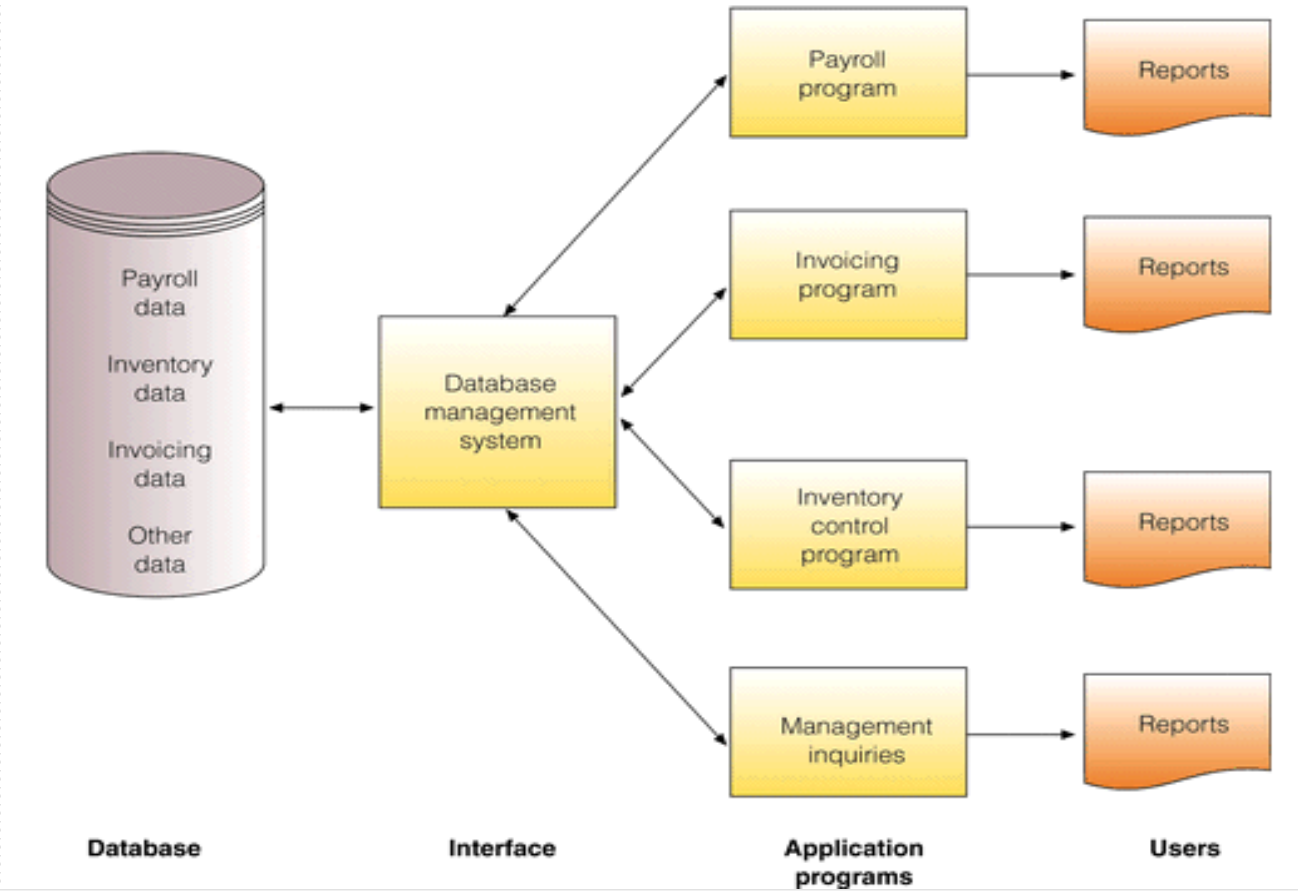


Database

- *"A collection of data, generally related to some subject or topic area and structured so as to allow relationships to be established between separate data items according to the various needs of users"*

G. Knott & N. Waites, Computing 3rd edition

Database Approach



Traditional Files vs. Databases: Pros and Cons

- Traditional File Advantages
  - Simplicity
  - Efficiency
  - Customization
- Database Advantages
  - Reduced data redundancy
  - Application/data independence
  - Better control
  - Flexibility
- Disadvantages of Database
  - High Cost
  - Need specialized staff
  - Increase vulnerability

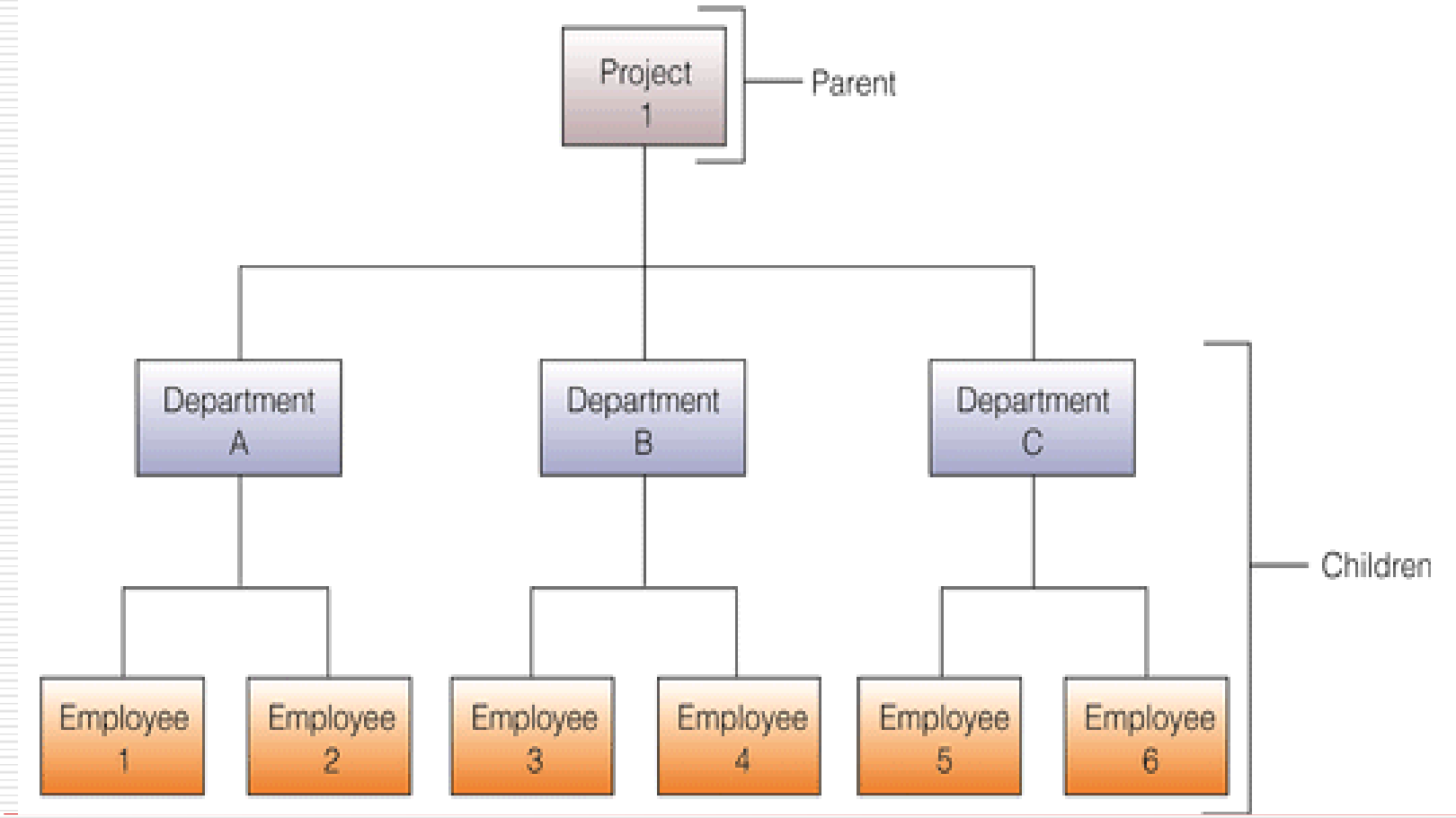
Database Management Systems (DBMS)

- Databases are usually managed by DBMS
- DBMS: A collection of programs that controls and prevents accidental or deliberate corruption of data by other application programs
- Functions of DBMS
  - Store, query and update data
  - Manage transaction
  - Control concurrent access to database
  - Maintain data consistency and integrity

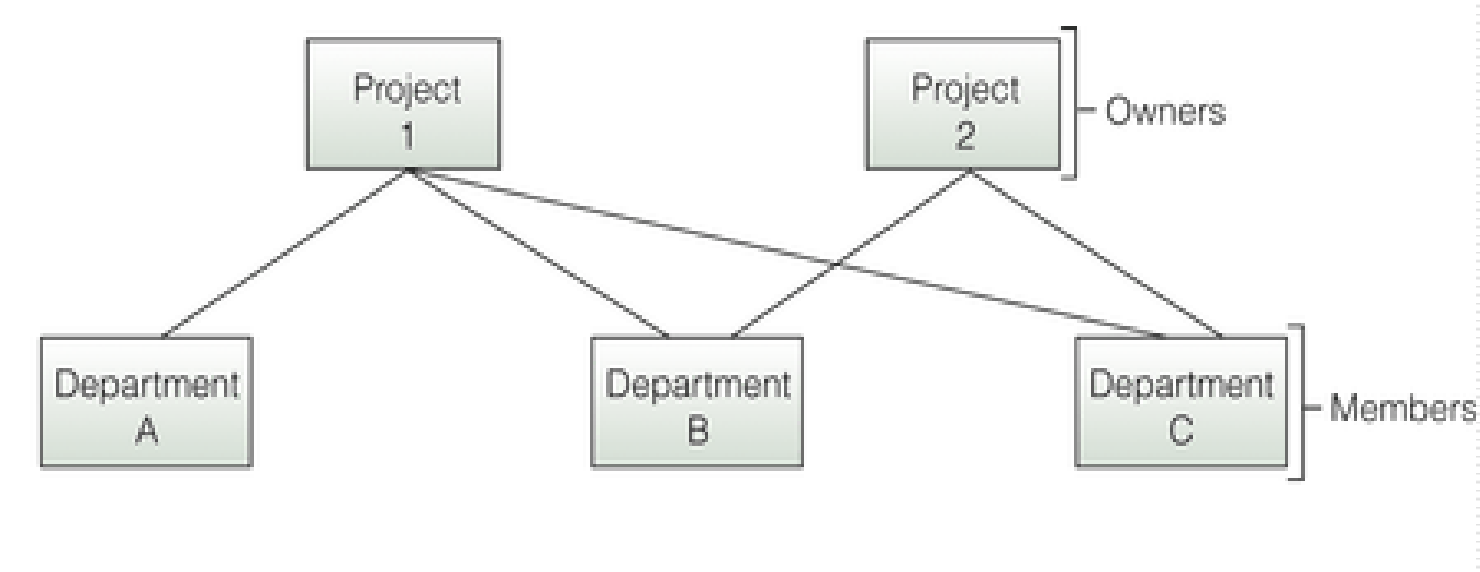
Types of Databases

- Hierarchical (tree)
- Network
- Relational

Hierarchical Database Model



Network Database Model



Relational Database Model

Data table 1: Project table

Project number	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales Manual	598

Data table 2: Department table

Dept. number	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

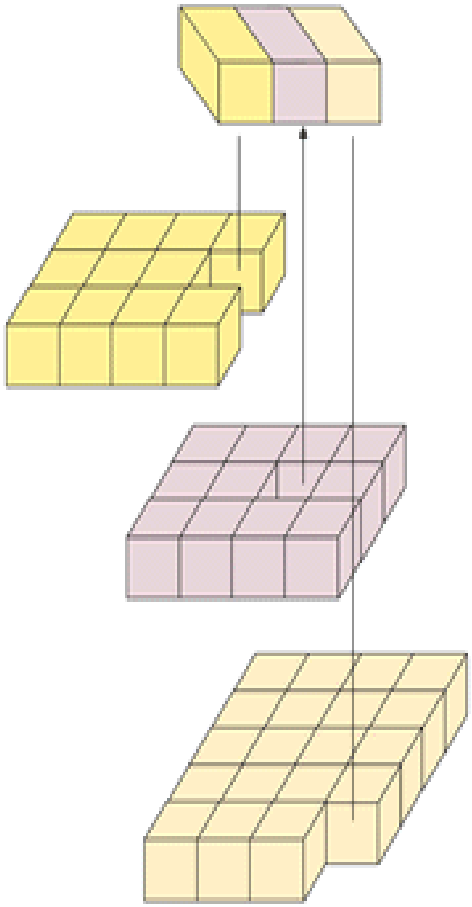
Data table 3: Manager table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-1997	257
549-77-1001	Buckley	Bill	02-17-1979	632
098-40-1370	Fiske	Steven	01-05-1985	598

Relational Models

- Describe data using a standard tabular format with all data elements placed in two-dimensional tables, called relations, that are the logical equivalent of files.
  - Domain
  - Selecting
  - Projecting
  - Joining

Linking Database Tables to Answer an Inquiry



Data table 1: Project table

Project number	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales Manual	598

Data table 2: Department table

Dept. number	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

Data table 3: Manager table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-1997	257
549-77-1001	Buckley	Bill	02-17-1979	632
098-40-1370	Fiske	Steven	01-05-1985	598

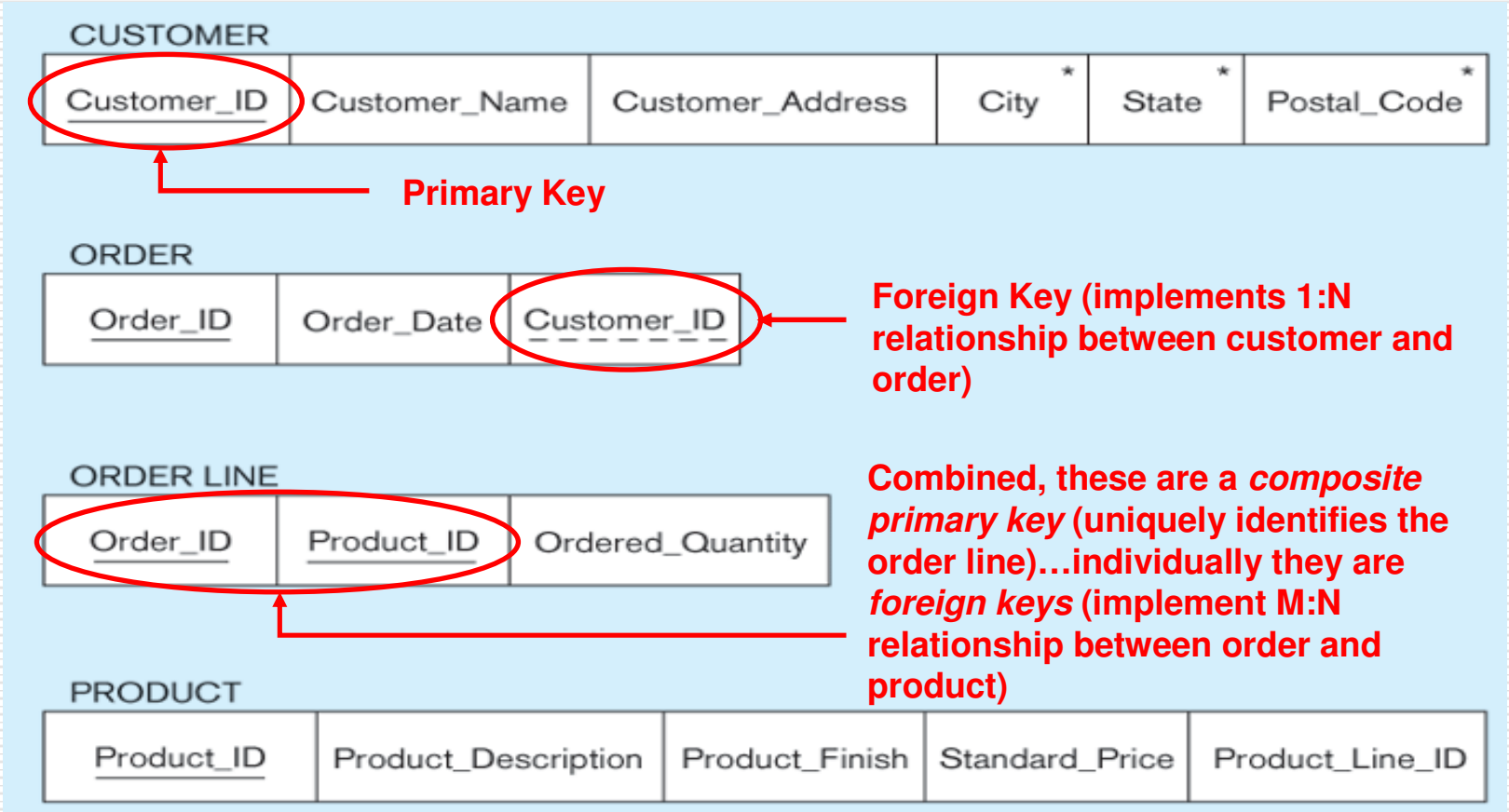
Keys

- Smallest set of attributes can be used to identify a row in a table
- Single key: composed of only one attribute
- Composite key: composed of two or more attributes
- A table can have many keys
- Primary key: a selected key
- Candidate keys: other keys, except primary key

Foreign Keys

- If a table uses the primary key of another table as attribute to refer to a row in the other table, the attribute is called referential key or foreign key.

Primary and Foreign Keys

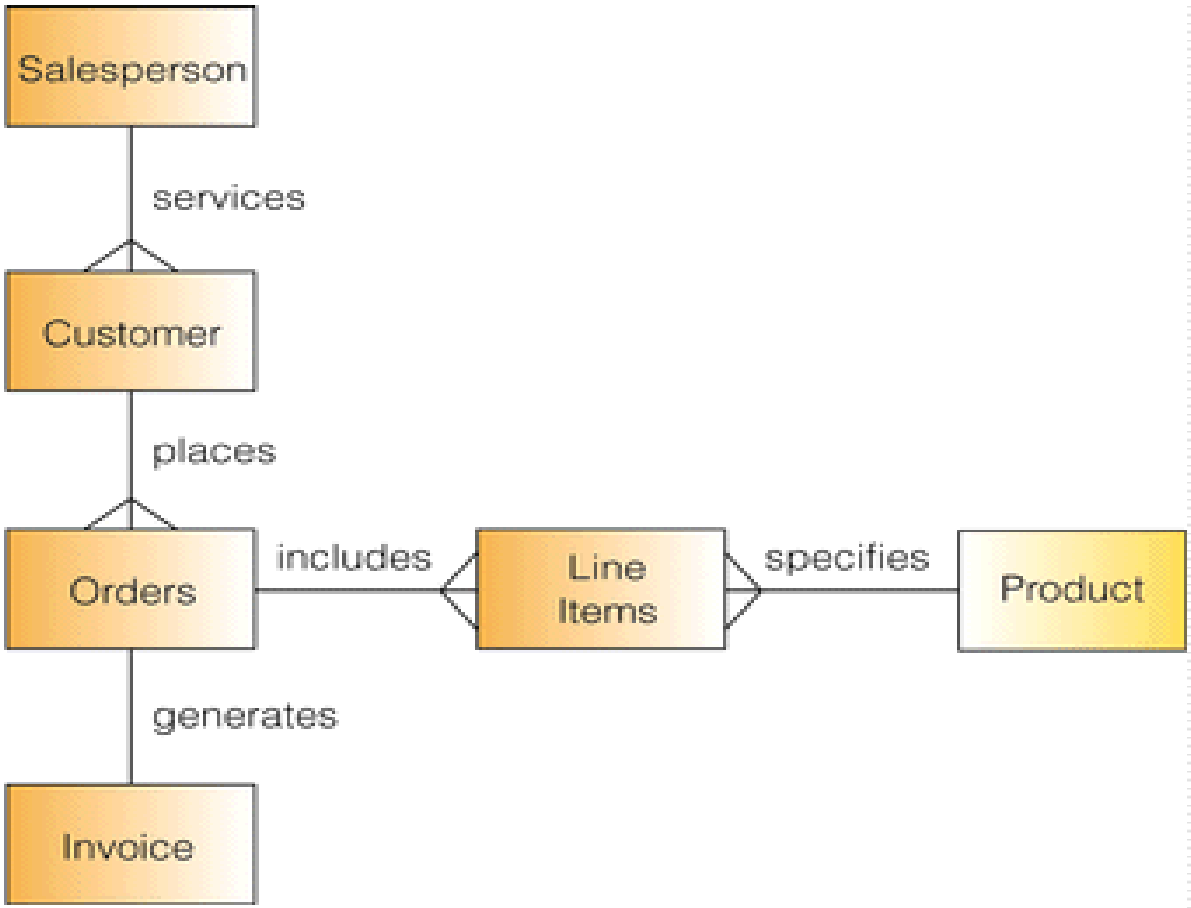


Database Design and Construction

Data Modelling

- To analyse and model the logical structures and relationships between data items
- To assist the data design process
- Entity Relationship Model is commonly used as a data modelling tool

Entity-Relationship Diagram for a Customer Ordering Database



Entity Relationship Model

- Entities
  - Objects of interest or relevant to the organisation, eg. Customer, stock
- Attributes
  - Properties of an entity which are identified as of interest
- Relationships
  - The relationship between entities
  - Relationship types:
    - One-to-one
    - One-to-many
    - Many-to-many

Database Design

- Convert entity relationship model to relational model
  - Convert entities into tables
  - Identify keys and attributes
  - Create referential keys (foreign keys) for relationships
- Decide data types and length for attributes

Data Normalisation

- A step-by-step process for analysing data into its constituent entities and attributes
- To improve database efficiency and to maintain data consistency and integrity
- Commonly, three normal forms are used:
  - First normal form (1NF)
  - Second normal form (2NF)
  - Third normal form (3NF)

Database Construction

- Use tools, eg. MS Access
- Use Structured Query Language (SQL)
  - Data Definition Language
    - for defining database and database tables
    - Commands: create database, create table, alter table, drop table, etc.
  - Data Manipulation Language
    - For manipulating the data within data
    - Commands: select, insert, update and delete

Distributed Databases

- In distributed database system, the database is subdivided over multiple computer systems but appear to users as a single database
- Considerations
  - Distribution transparency
  - Transaction transparency
    - Integrity must be maintained across databases

Pros and Cons

- Advantages
  - Reflect organisational structure
  - Local autonomy
  - Improve availability and performance
  - Modularity
- Disadvantages
  - Complexity
  - Security
  - Difficult to maintain integrity