SYSTEM DESIGN:

DATA DESIGN

ITSRAQA Group 5 | Dionisio, Manzano, Ostia, Salvador, Uy

OPENING PRAYER

DATA DESIGN CONCEPTS 01.

DBMS COMPONENTS 02.

WEB-BASED DESIGN 03.

DATA DESIGN TERMS 04.

ENTITY-RELATIONSHIP DIAGRAMS 05.

06. DATA NORMALIZATION

07. USING CODES

08. DATA STORAGE & ACCESS

09. DATA CONTROL

10. SUMMARY

01.

DATA DESIGN CONCEPTS

WHAT IS A DATA STRUCTURE?

A framework for organizing, storing, and managing data. Data structures consist of either files or tables that interact in various ways.

LEGACY SYSTEMS: FILE ORIENTED

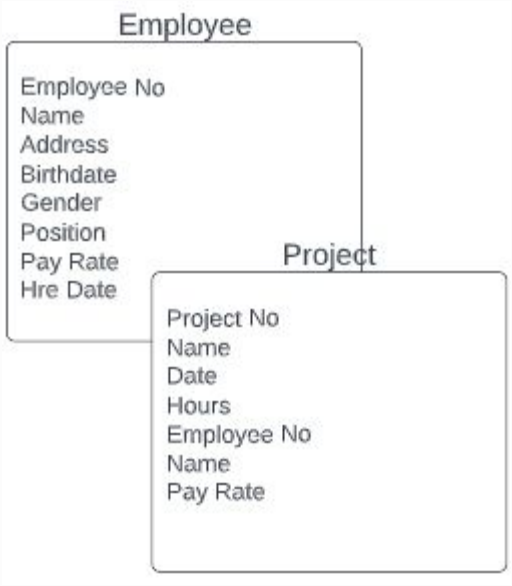
DATA DESIGN EXAMPLES

FILE-ORIENTED SYSTEMS

RELATIONAL DATABASE / MODEL

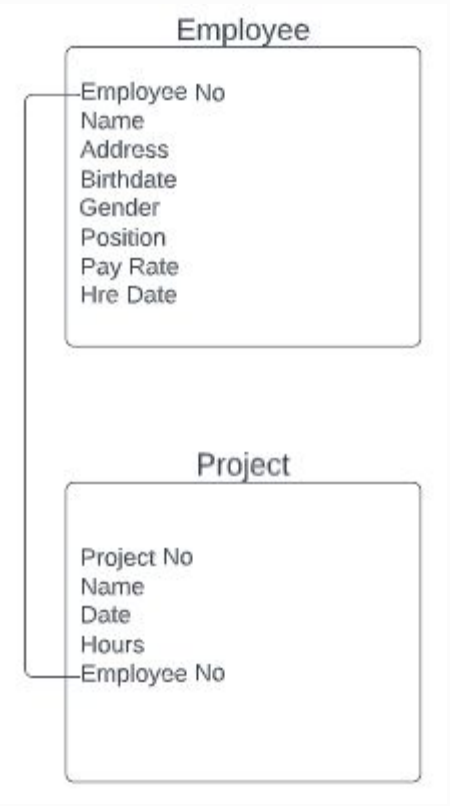
FILE-ORIENTED SYSTEMS

- Two separate and independent files
- Some data entries can be stored in two independent places

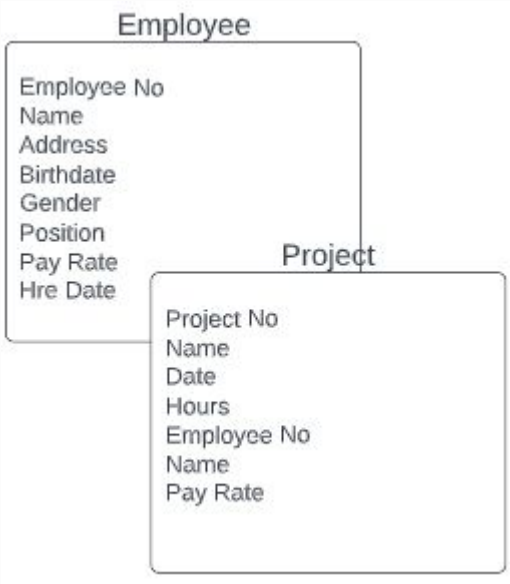


RELATIONAL DATABASE OR MODEL

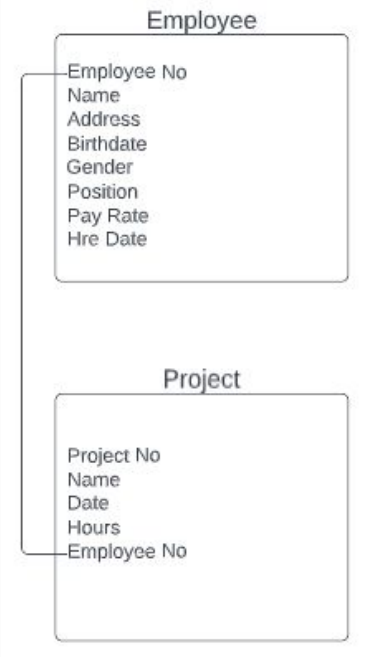
- In database management systems (DBMSs), the only duplicate entry is a common field that connects tables together.



DATA DESIGN EXAMPLES



FILE-ORIENTED SYSTEMS



RELATIONAL DATABASE / MODEL

IS FILE PROCESSING STILL IMPORTANT?

- File processing is a fairly dated approach.
- Method to handle large volumes of structured data on a regular basis.
- Many older legacy systems utilized file processing because it worked well with mainframe hardware and batch input.
- Although it has very limited use today, file processing can be cost-effective in certain situations.

THE DATABASE ENVIRONMENT: ADVANTAGES

SCALABILITY

a system can be expanded, modified, or downsized

ECONOMY OF SCALE

Database design allows better utilization of hardware

ENTERPRISE-WIDE APPLICATION

A DBMS is typically managed by a person called a database administrator (DBA)

STRONGER STANDARDS

standards for data names, formats, and documentation are followed uniformly

BETTER SECURITY

procedures to ensure that only legitimate users can access the database; have different levels of access

DATA INDEPENDENCE

Systems that interact with a DBMS are relatively independent of how the physical data is maintained.

02.

DBMS COMPONENTS

INTERFACES

USERS

Work with predefined queries and switchboard commands; also use query languages to access stored data

DATABASE ADMINISTRATORS

Also known as a DBA; is responsible for DBMS management and support.

RELATED INFORMATION SYSTEMS

A DBMS can support several related information systems that provide input to, and require specific data from, the DBMS

INTERFACES

USERS

Work with predefined queries and switchboard commands; also use query languages to access stored data

INTERFACES

USERS

Work with predefined queries and switchboard commands; also use query languages to access stored data

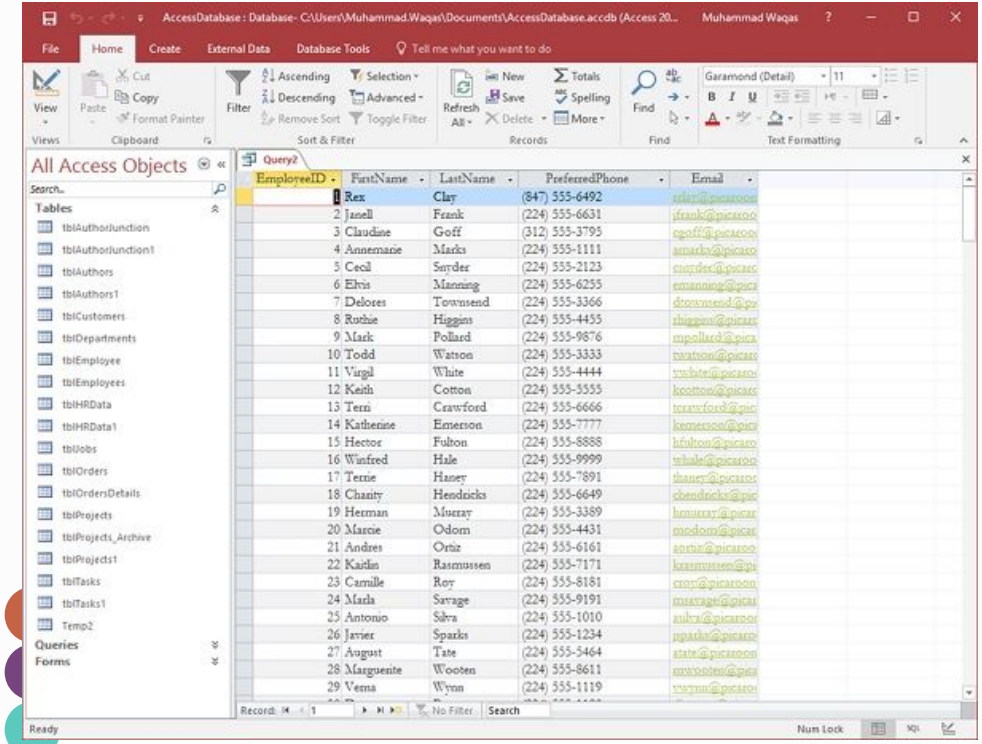
DATABASE ADMINISTRATORS

Also known as a DBA; is responsible for DBMS management and support.

RELATED INFORMATION SYSTEMS

A DBMS can support several related information systems that provide input to, and require specific data from, the DBMS

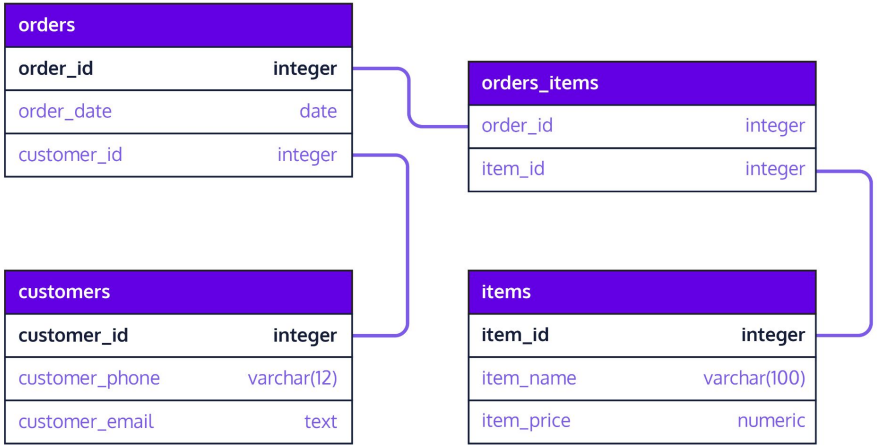
DATA MANIPULATION LANGUAGE



A data manipulation language (DML) controls database operations, including storing, retrieving, updating, and deleting data.

SCHEMA

- The complete definition of a database, including descriptions of all fields, tables, and relationships, is called a schema.
- Subschema - is a view of the database used by one or more systems or users. A subschema defines only those portions of the database that a particular system or user needs or is allowed to access.

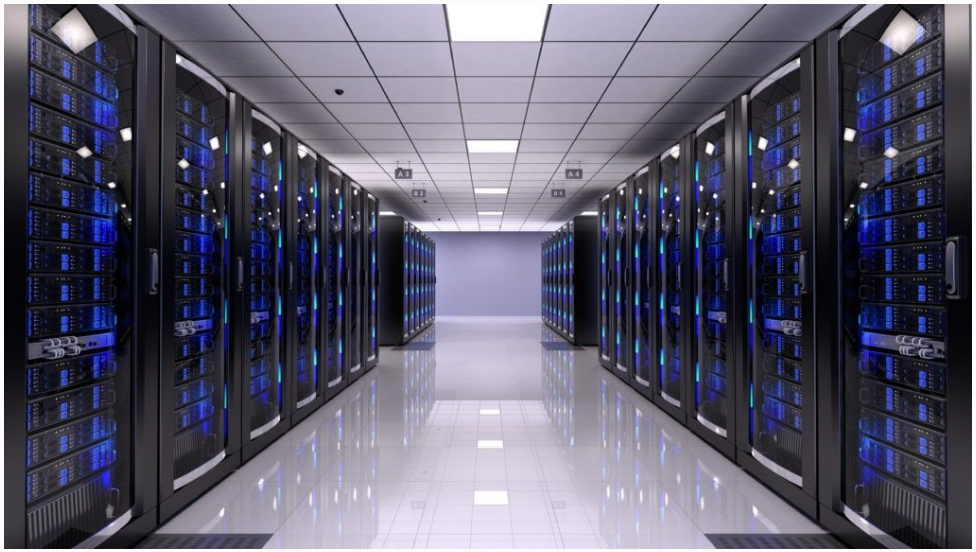


PHYSICAL DATA REPOSITORY

The physical repository might be centralized, or it might be distributed at several locations. In addition, the stored data might be managed by a single DBMS or several systems.

MICROSOFT OPEN DATABASE CONNECTIVITY (ODBC)

- Makes it possible for applications to access data from a variety of database management systems (DBMSs).
- ODBC is a low-level, high-performance interface that is designed specifically for relational data stores.



03.

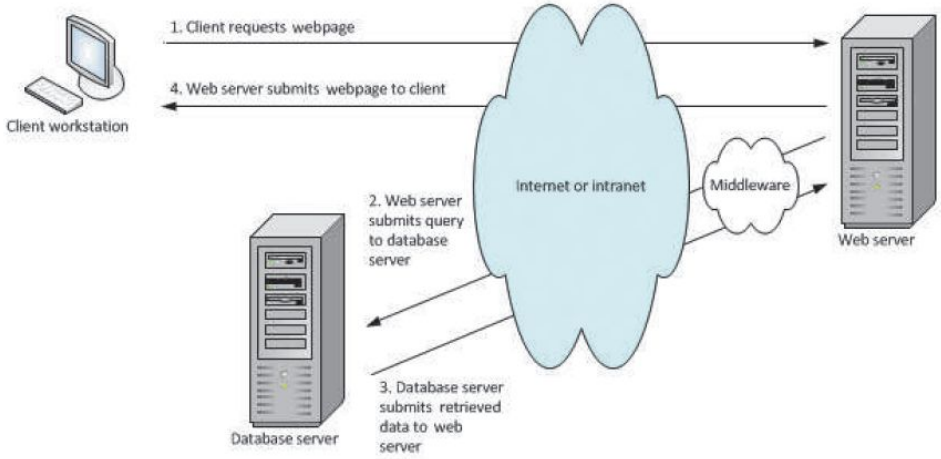
WEB-BASED DESIGN

WEB-BASED DESIGN

Major characteristics of web-based design

Global access	The internet enables worldwide access.
Ease of use	Web browsers provide a familiar interface that is user-friendly and easily learned.
Multiple platforms	Web-based design is not dependent on a specific combination of hardware or software.
Cost effectiveness	Flexibility is high because numerous outsourcing options exist for development, hosting, maintenance, and system support.
Security issues	A combination of good design, software that can protect the system and detect intrusion, stringent rules for passwords and user identification,
Adaptability issues	Migrating a traditional database design to the web can require design modification, additional software, and some added expense.

CONNECTING TO THE WEB



To access data in a web-based system, the database must be connected to the Internet or intranet. The database and the Internet speak two different languages, however. The objective is to connect the database to the web and enable data to be viewed and updated.

04.

DATA DESIGN TERMS

DEFINITIONS

ENTITY

a person, place, thing, or event for which data is collected and maintained

TABLE OR FILE

contains a set of related records that store data about a specific entity

FIELD

is a single characteristic or fact about an entity

RECORD

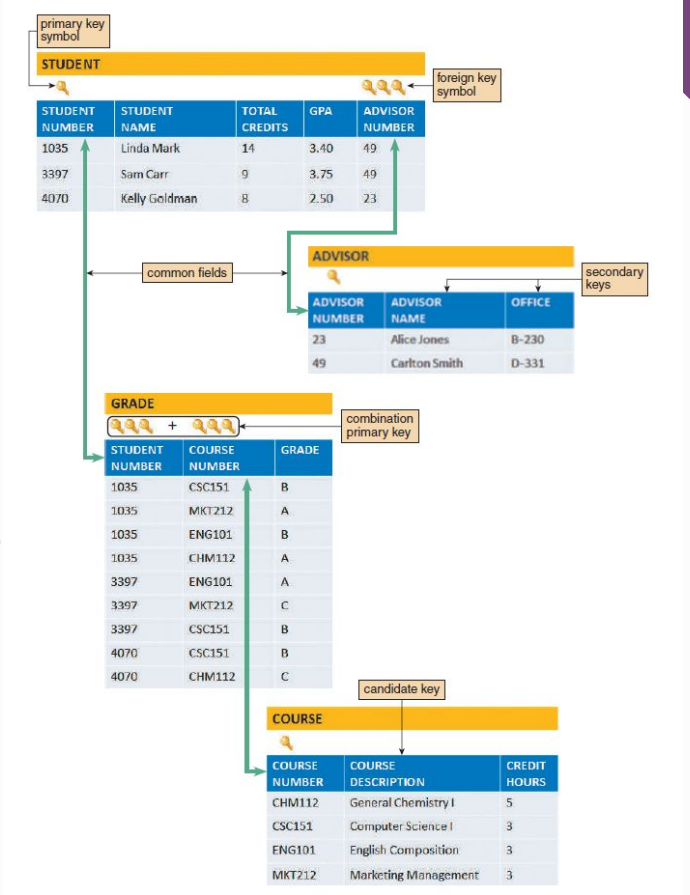
is a set of related fields that describes one instance, or occurrence, of an entity

KEY FIELDS

PRIMARY KEY

A primary key is a field or combination of fields that uniquely and minimally identifies a particular member of an entity.

- **Combination Key/Composite Key** - a combination of two or more columns in a table

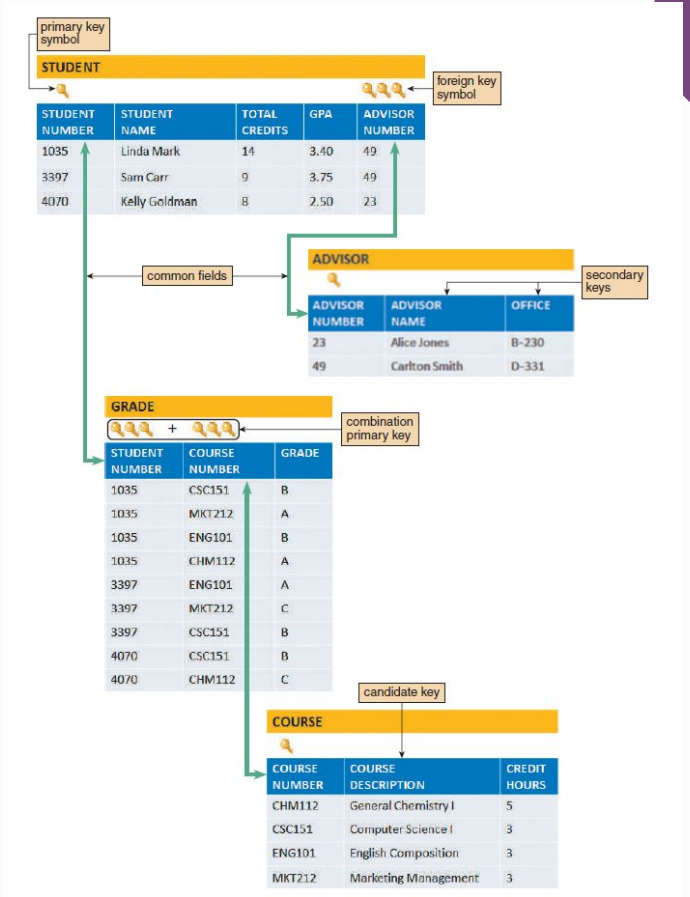


KEY FIELDS

CANDIDATE KEY

Any field that can serve as a primary key is called a candidate key.

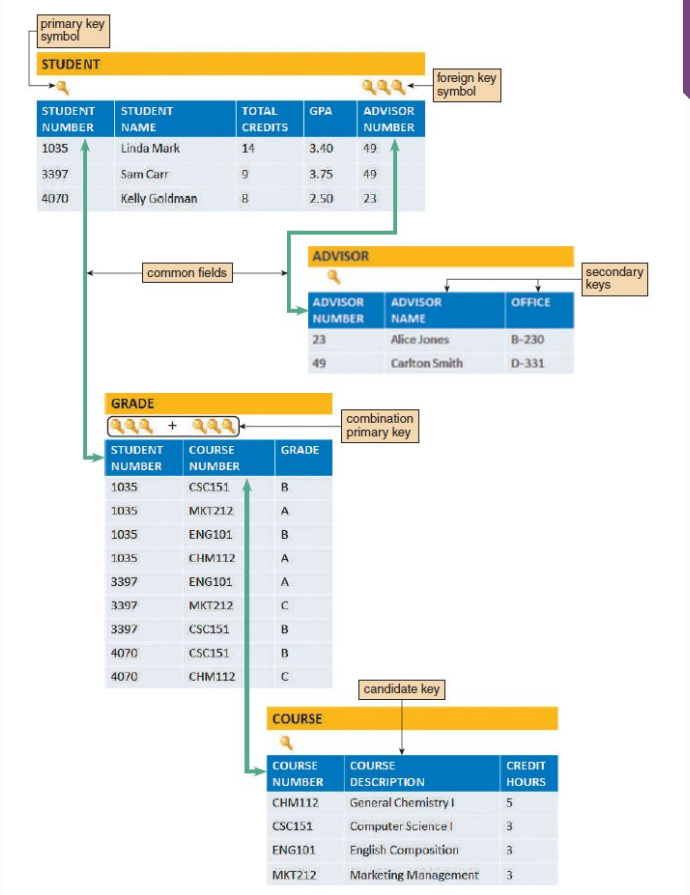
- **Nonkey Field** - any field that is not a primary key or a candidate key



KEY FIELDS

FOREIGN KEY

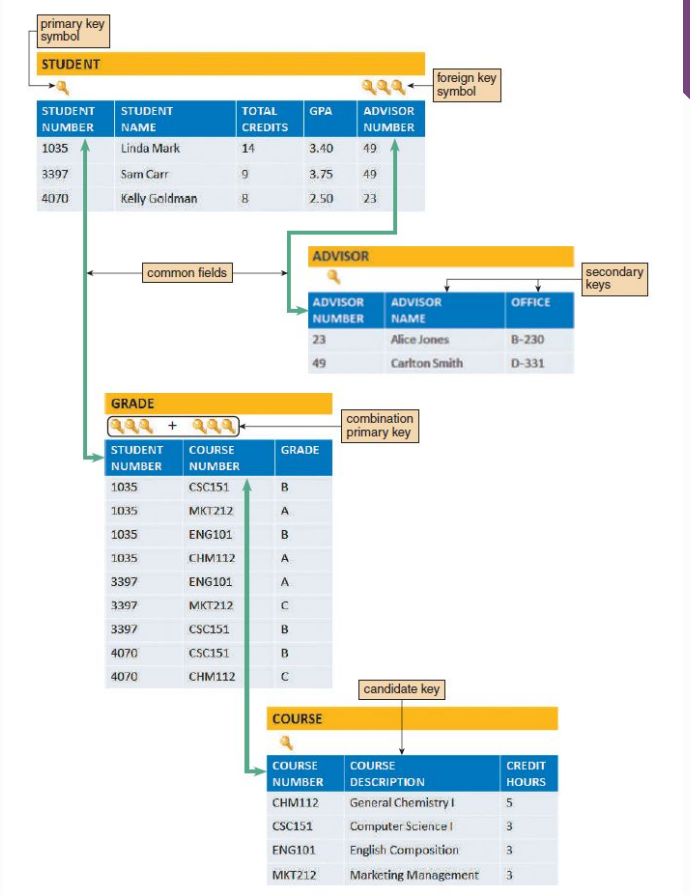
A foreign key is a field in one table that must match a primary key value in another table in order to establish the relationship between the two tables.



KEY FIELDS

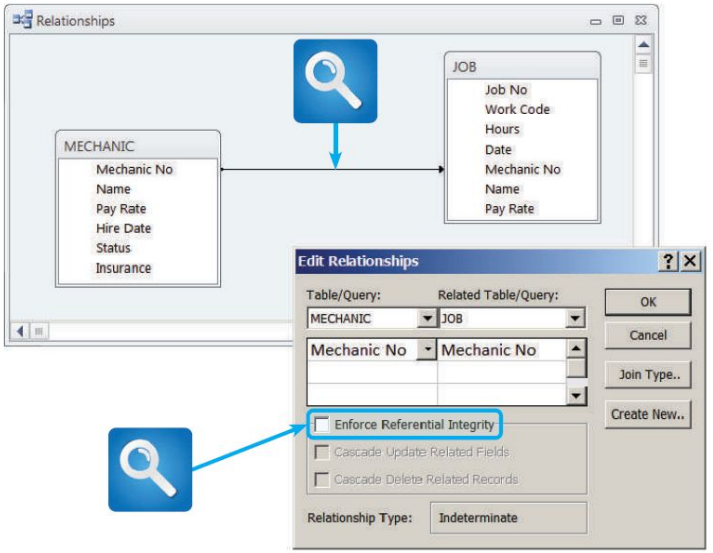
SECONDARY KEY

A secondary key is a field or combination of fields that can be used to access or retrieve records.



REFERENTIAL INTEGRITY

- Validity checks can help avoid data input errors. One type of validity check, called **referential integrity**, is a set of rules that avoids data inconsistency and quality problems.
- a foreign key value cannot be entered in one table unless it matches an existing primary key in another table.
- can also prevent the deletion of a record if the record has a primary key that matches foreign keys in another table.



05.

ENTITY-RELATIONSHIP DIAGRAMS

ENTITY

a person, place, thing, or event for which data is collected and maintained.

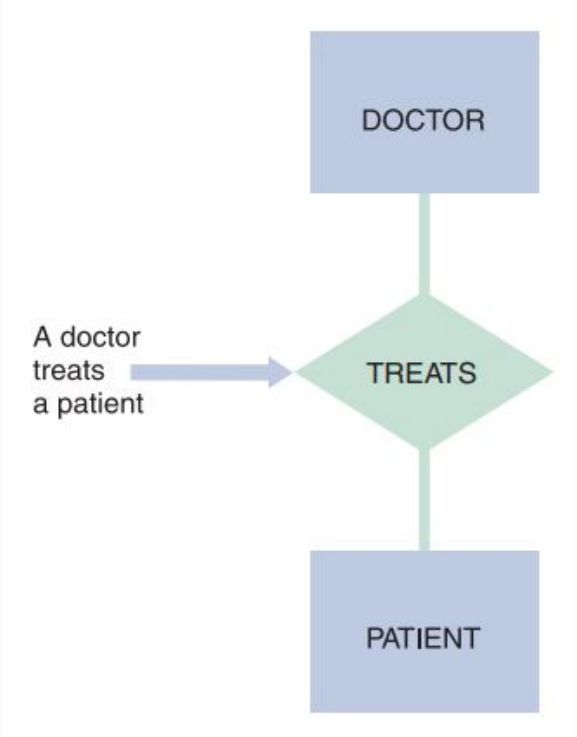
- Customers
- Sales Regions
- Products
- Orders

ENTITY-RELATIONSHIP DIAGRAM (ERD)

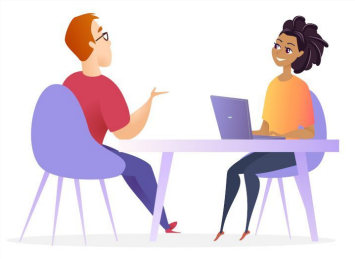
- shows the logical relationships and interactions among system entities.
- provides an overall view of the system and a blueprint for creating the physical data structures.

DRAWING AN ERD

1. list the **entities**
2. represent entities as **rectangles** and relationships as **diamond** shapes.
3. The entity rectangles are labeled with **singular nouns**, and the relationship diamonds are labeled with **verbs**,
4. Present in a **top-to-bottom** or **left-to-right** fashion.



3 TYPES OF RELATIONSHIP



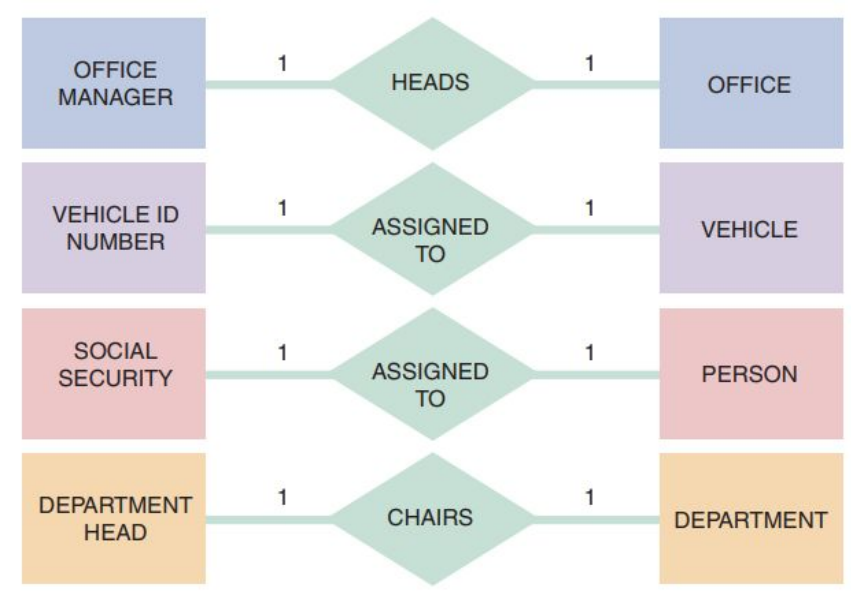
ONE-TO-ONE

ONE-TO-MANY

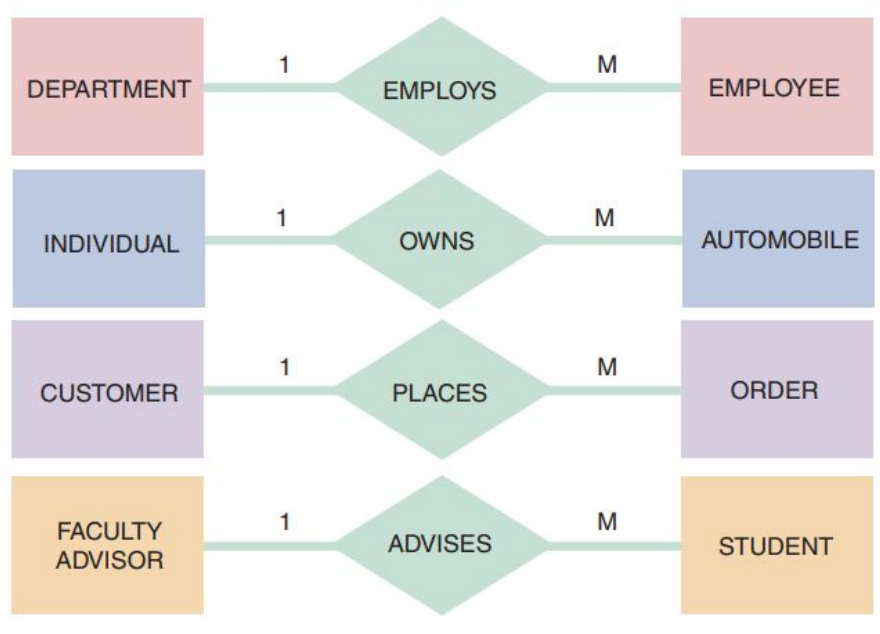
MANY-TO-MANY

ONE-TO-ONE RELATIONSHIP (1:1)

Exists when **exactly one of the second entity** occurs for **each instance** of the first entity.



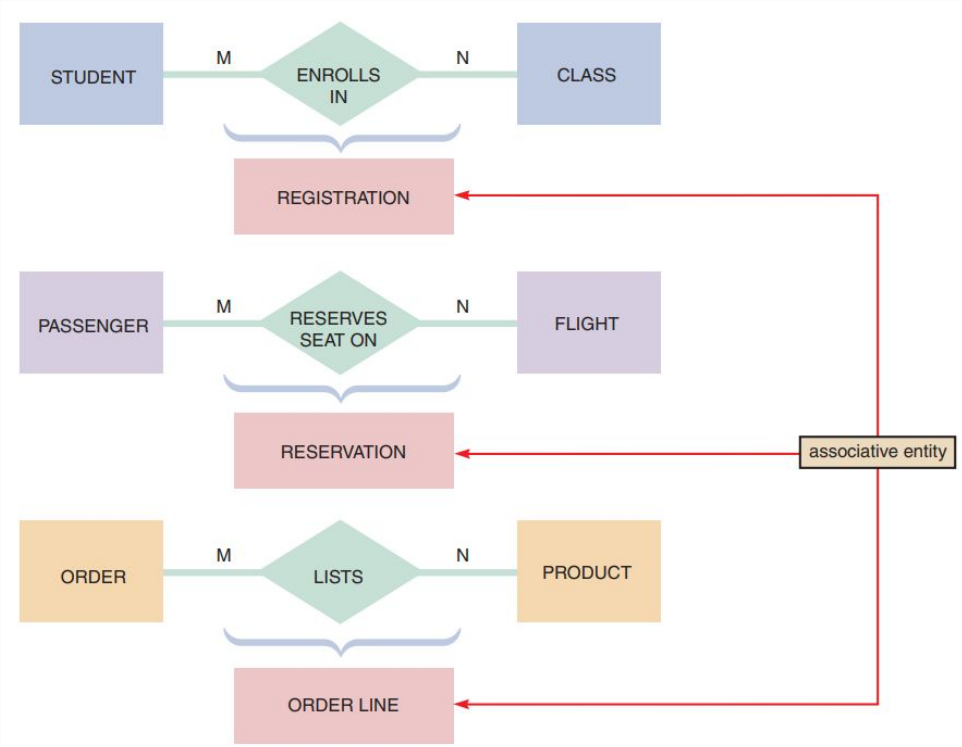
ONE-TO-MANY RELATIONSHIP (1:M)



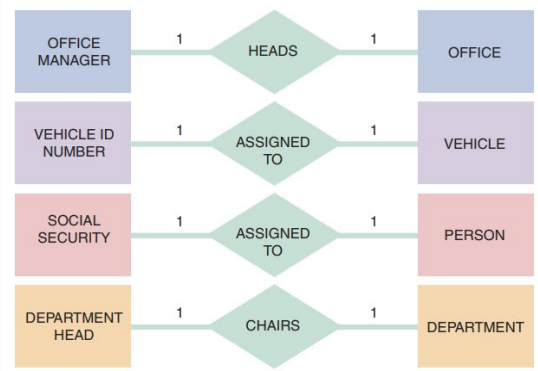
Exists when **one** occurrence of the first entity can relate to **many** instances of the second entity, but each instance of the second entity can associate with only one instance of the first entity

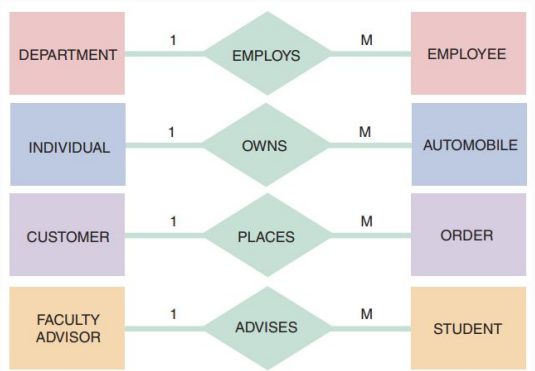
MANY-TO-MANY RELATIONSHIP (M:N)

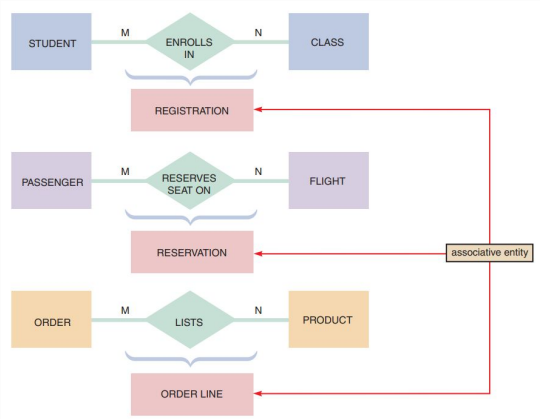
Exists when **one instance** of the first entity can relate to **many instances** of the second entity, and **one instance** of the second entity can relate to **many instances** of the first entity.



ENTITY-RELATIONSHIP DIAGRAM



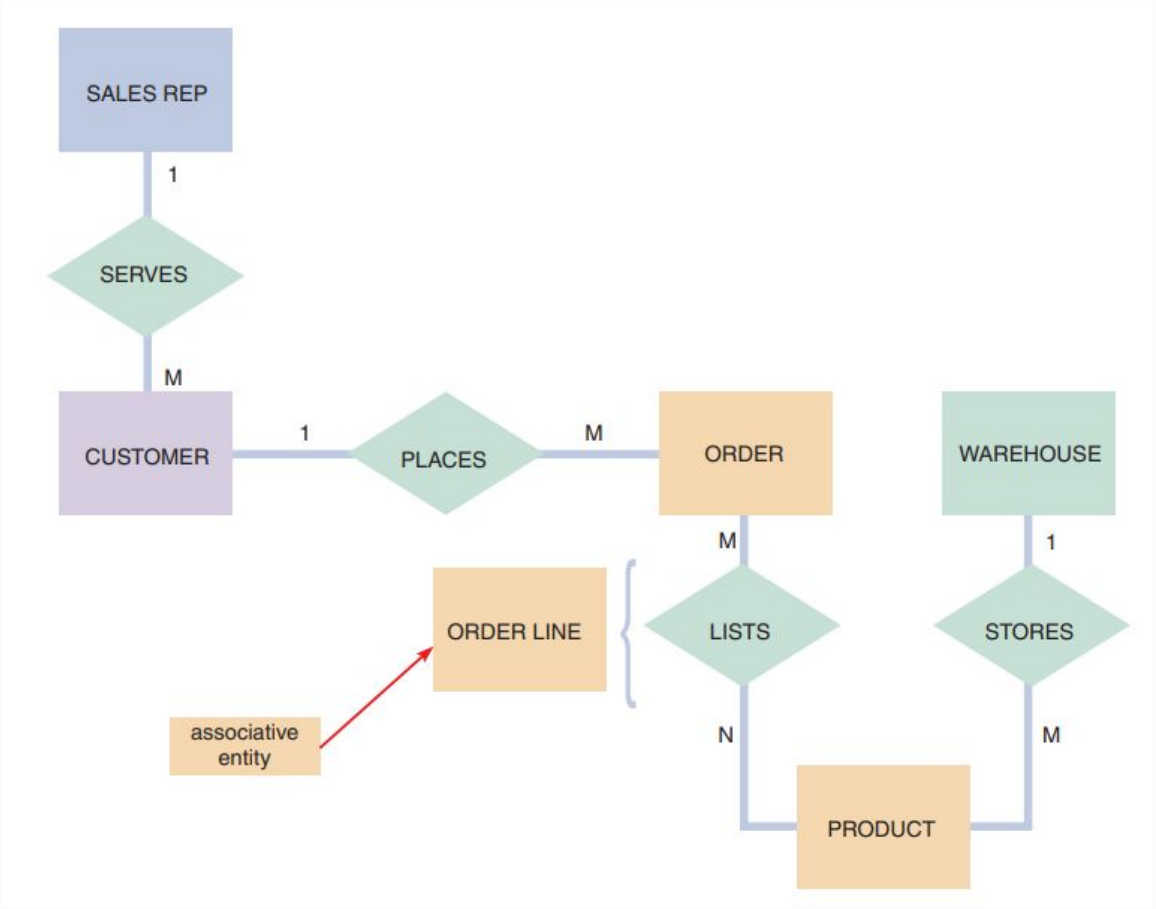




Notice that an M:N relationship is different from 1:1 or 1:M relationships

because the event or transaction that links the two entities is actually a third entity, called an **associative entity**, that has its own characteristics.

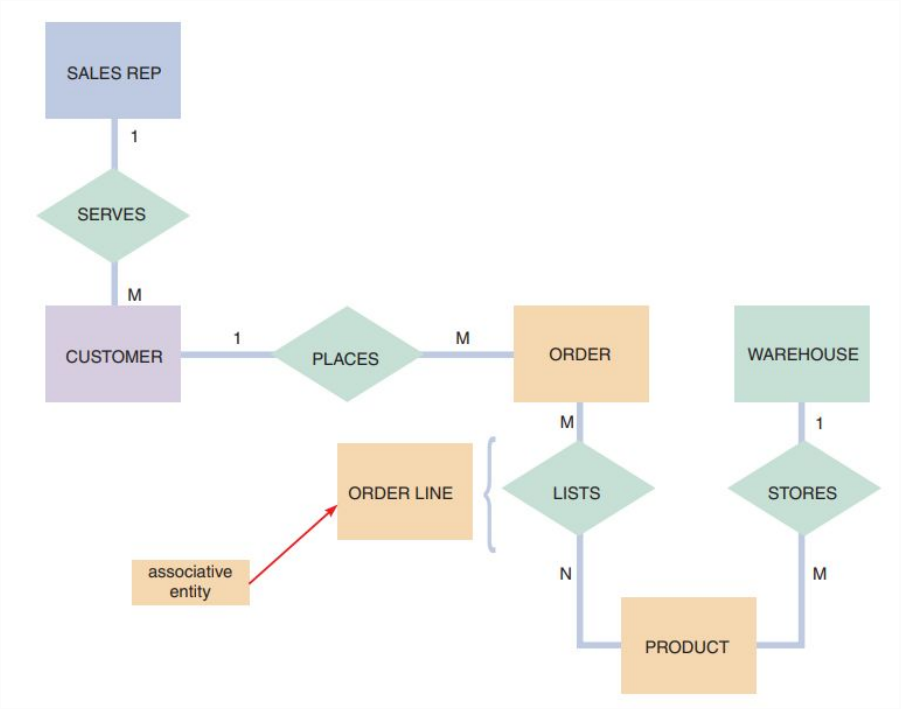
ENTITY-RELATIONSHIP DIAGRAM



CARDINALITY

Cardinality

- the numeric relationship between two entities and shows how instances of one entity relate to instances of another entity.



CARDINALITY

common method of cardinality notation is called **crow's foot notation**

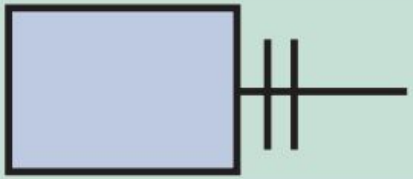
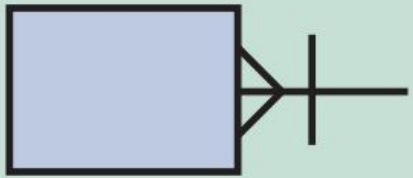
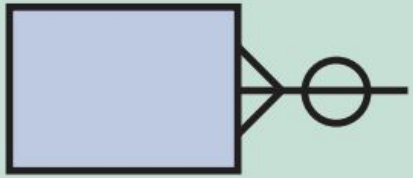
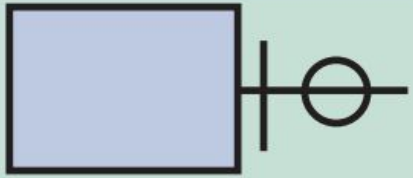
ONE

ONE AND ONLY ONE

ZERO

MANY

ENTITY-RELATIONSHIP DIAGRAM

SYMBOL	MEANING	UML REPRESENTATION
	One and only one	1
	One or many	1..*
	Zero, or one, or many	0..*
	Zero, or one	0..1

EXAMPLE

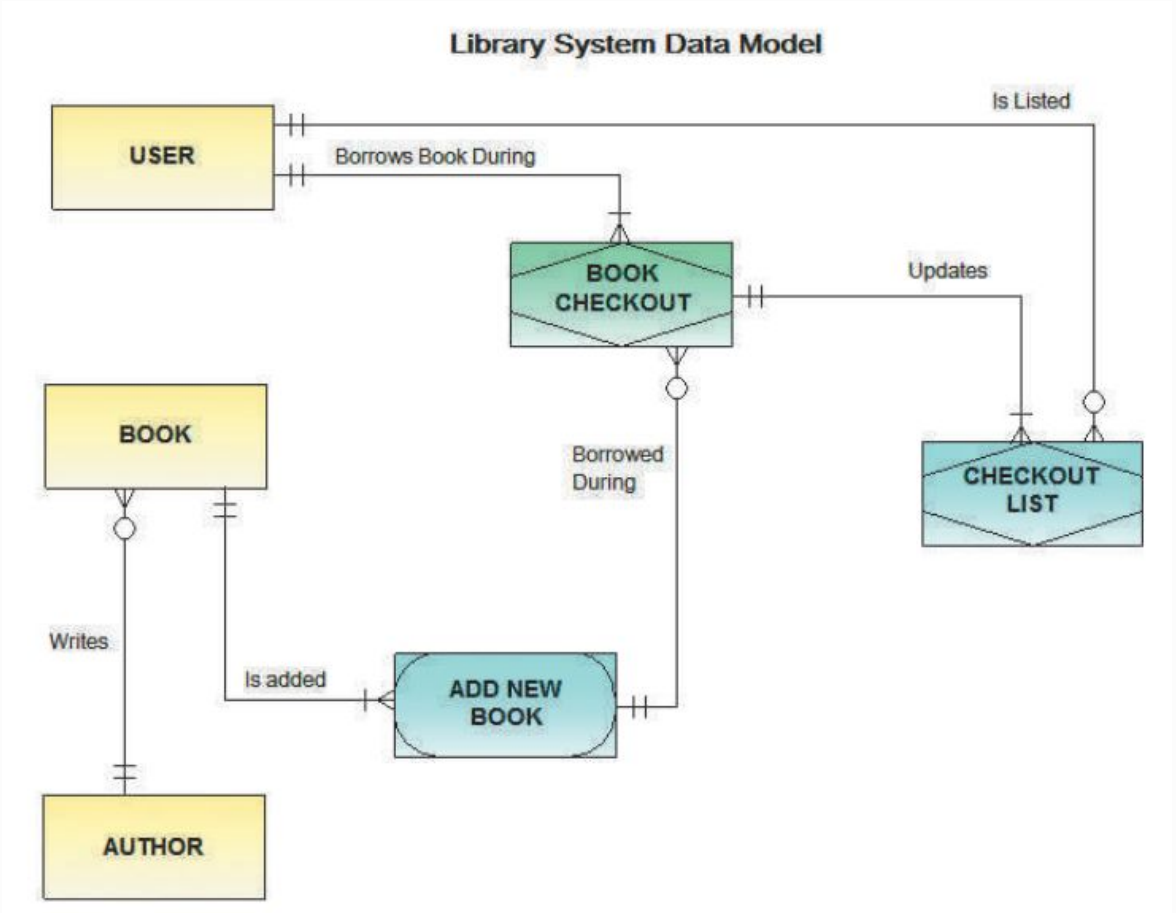
One and only one CUSTOMER can place anywhere from zero to many of the ORDER entity.

One and only one ORDER can include one ITEM ORDERED or many.

One and only one EMPLOYEE can have one SPOUSE or NONE.

One EMPLOYEE, or many employees, or none, can be assigned to one PROJECT, or many projects, or none.

EXAMPLE


Library System Data Model
USER
Is Listed
Borrows Book During
BOOK CHECKOUT
Updates
BOOK
Borrowed During
CHECKOUT LIST
Writes
Is added
ADD NEW BOOK
AUTHOR

06.

DATA NORMALIZATION

NORMALIZATION

Based on the work of **Edgar Codd**

It is the process of creating **table designs** by assigning specific fields or attributes to each table in the database.

Involves applying a set of rules that help identify and correct inherent problems and complexities in **table designs**.

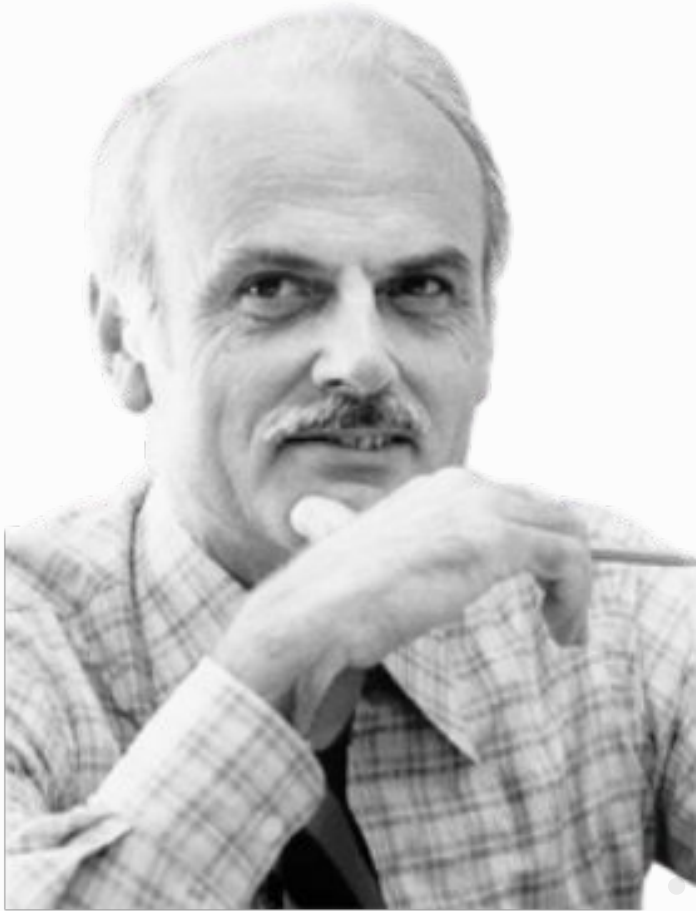


TABLE DESIGN

Specifies the fields and identifies the primary key in a particular table or file.

Working with a set of initial table designs, normalization is used to develop an overall database design that is simple, flexible, and free of data redundancy.

4 STAGES OF NORMALIZATION PROCESS

UNNORMALIZED DESIGN

FIRST NORMAL FORM

SECOND NORMAL FORM

THIRD NORMAL FORM

constitute a progression in which the third normal form represents the best design

STANDARD NOTATION FORMAT

The standard notation format in the following examples starts with the

- name of the table,
- parenthetical expression that contains the field names separated by commas.

EXAMPLE:

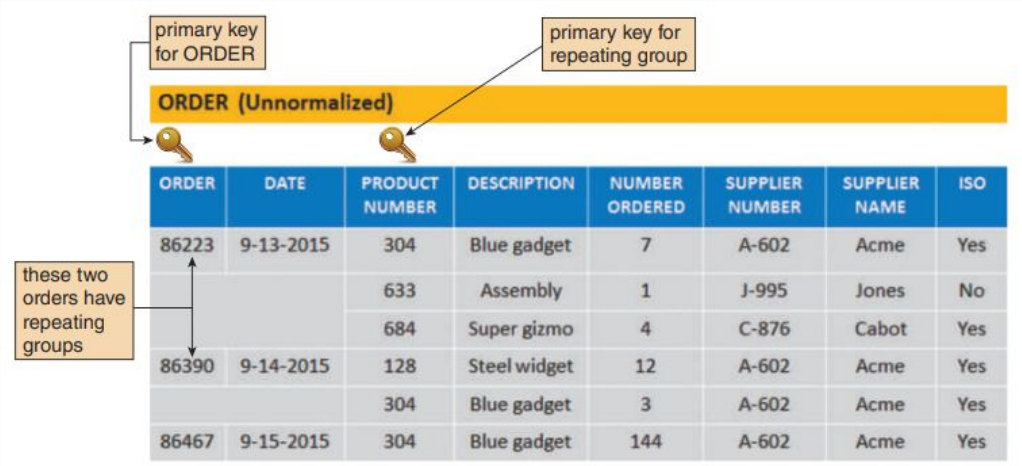
Primary Key

NAME (**FIELD 1**, FIELD 2, FIELD 3)

STANDARD NOTATION FORMAT

During data design, the analyst must be able to recognize a repeating group of fields.

A **repeating group** is a set of one or more fields that can occur any number of times in a single record, with each occurrence having different values.



ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
		633	Assembly	1	J-995	Jones	No
		684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
		304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

UNNORMALIZED DESIGN

A table design that contains a **repeating group**.

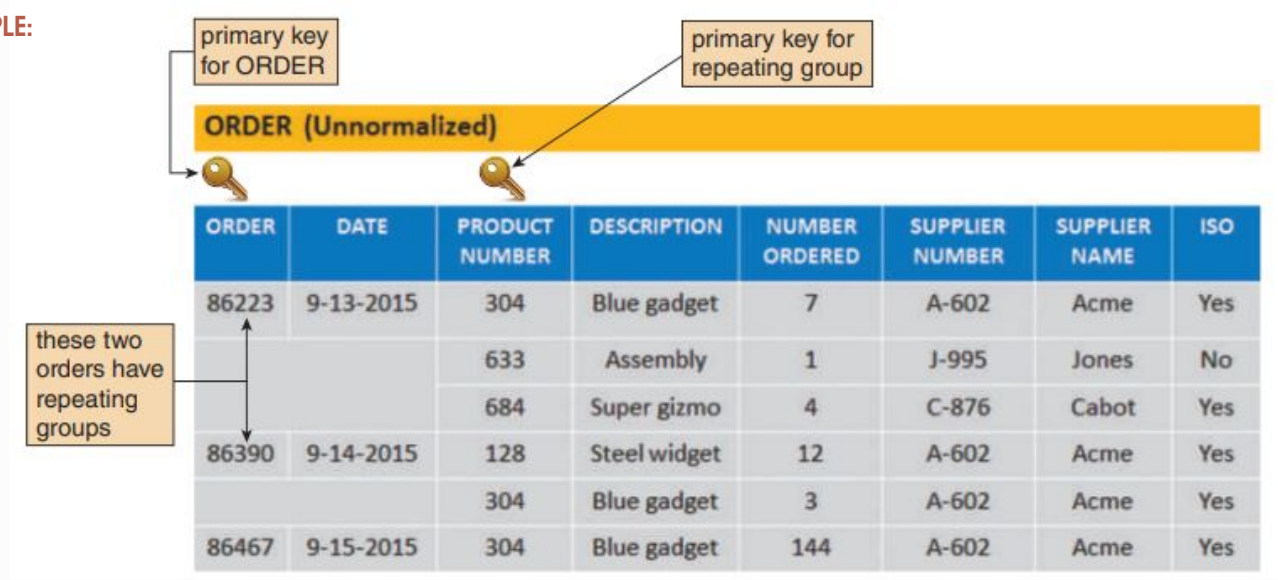
The standard notation method for representing an **unnormalized design** is to enclose the repeating group of fields within a second set of parentheses. An example of an unnormalized table looks like this:

FORMAT: Primary Key

NAME (**FIELD 1**, FIELD 2 (REPEATING FIELD 1, REPEATING FIELD 2

UNNORMALIZED DESIGN

EXAMPLE:



ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
		633	Assembly	1	J-995	Jones	No
		684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
		304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

ORDER (**ORDER**, DATE, (**PRODUCT NUMBER**, DESCRIPTION, NUMBER ORDERED, SUPPLIER NUMBER, SUPPLIER NAME, ISO))

FIRST NORMAL FORM

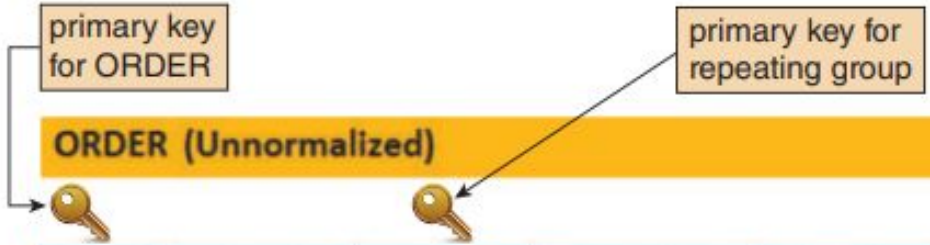
A table is in First Normal Form (1NF) if it **does not contain a repeating group**.

To convert an unnormalized design to 1NF, the table's primary key must be expanded to include the primary key of the repeating group.

EXAMPLE: Primary Key

NAME (**FIELD 1**, FIELD 2, **FIELD 3**, FIELD 4)

FIRST NORMAL FORM

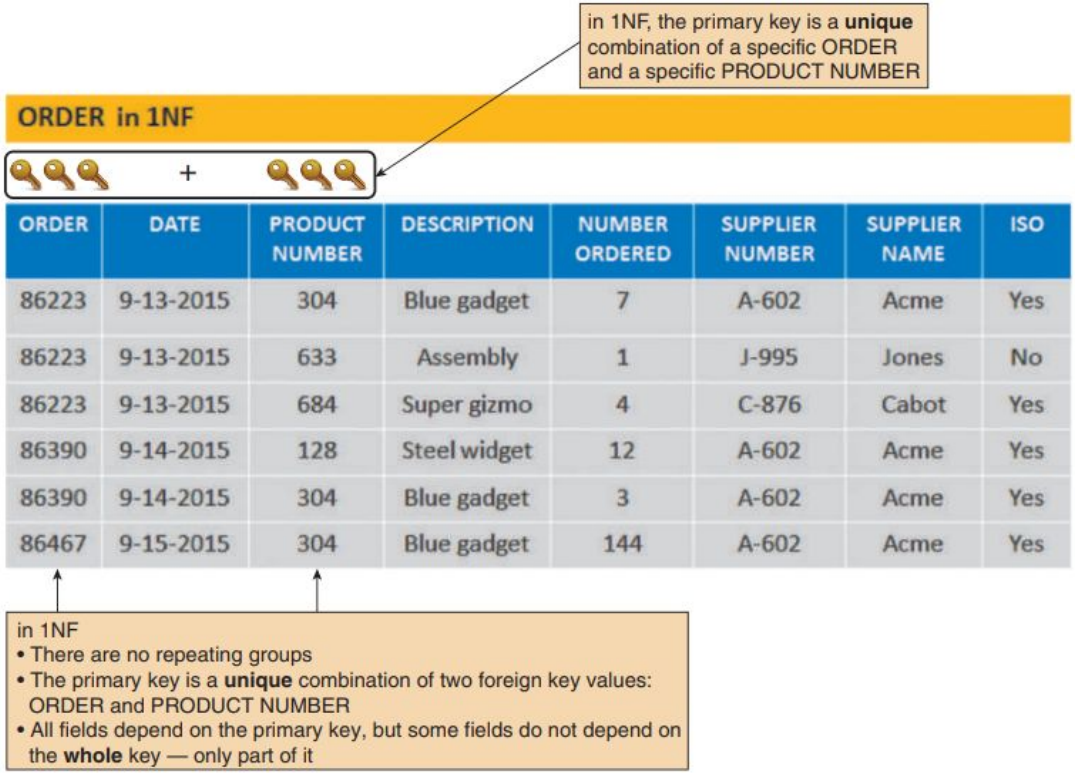


ORDER (Unnormalized)

ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
		633	Assembly	1	J-995	Jones	No
		684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
		304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

these two orders have repeating groups

FIRST NORMAL FORM



ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
86223	9-13-2015	633	Assembly	1	J-995	Jones	No
86223	9-13-2015	684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
86390	9-14-2015	304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

ORDER (ORDER, DATE, PRODUCT NUMBER, DESCRIPTION, NUMBER ORDERED, SUPPLIER NUMBER, SUPPLIER NAME, ISO)

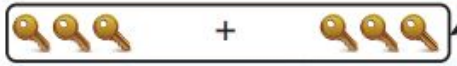
SECOND NORMAL FORM

A table design is in **second normal form (2NF)** if

- it is in 1NF and
- Fields that are not part of the primary key are **functionally dependent** on the entire primary key.

1NF design that has a primary key that consists of only one field, the problem of partial dependence does not arise and is automatically in 2NF

ORDER in 1NF



ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
86223	9-13-2015	633	Assembly	1	J-995	Jones	No
86223	9-13-2015	684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
86390	9-14-2015	304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

ORDER (ORDER, DATE, PRODUCT NUMBER, DESCRIPTION, NUMBER ORDERED, SUPPLIER NUMBER, SUPPLIER NAME, ISO)

CONVERTING 1NF TO 2NF

STEP 1

Create and name a separate table for each field in the existing primary key.

ORDER in 2NF

ORDER	ORDER DATE
86223	9-13-2015
86390	9-14-2015
86467	9-15-2015

PRODUCT in 2NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER	SUPPLIER NAME	ISO
128	Steel widget	A-602	Acme	Yes
304	Blue gadget	A-602	Acme	Yes
633	Assembly	J-995	Jones	No
684	Super gizmo	C-876	Cabot	Yes

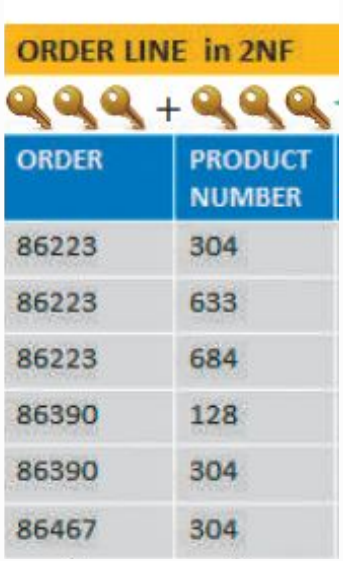
ORDER (ORDER, . . .)

PRODUCT (PRODUCT NUMBER, . . .)

CONVERTING 1NF TO 2NF

STEP 2

Create a new table for each possible combination of the original primary key fields.



ORDER LINE in 2NF

ORDER	PRODUCT NUMBER
86223	304
86223	633
86223	684
86390	128
86390	304
86467	304

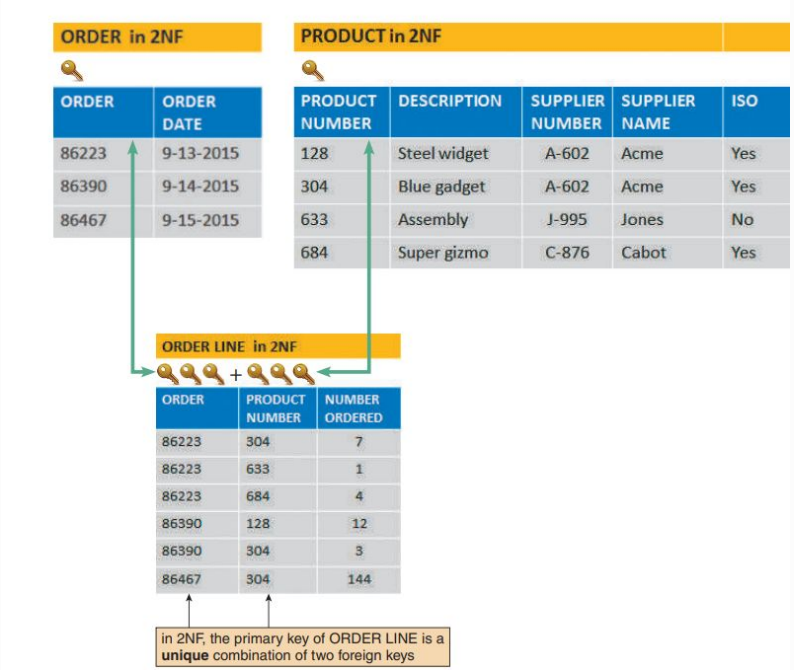
ORDER LINE (ORDER, PRODUCT NUMBER)

CONVERTING 1NF TO 2NF

STEP 3

Place each field with its appropriate primary key, which is the minimal key on which it functionally depends.

Remove any table that did not have any additional fields assigned to it. The remaining tables are the 2NF version of the original table.



The three tables can be shown as:

ORDER (ORDER, DATE)

PRODUCT (PRODUCT NUMBER, DESCRIPTION, SUPPLIER NUMBER, SUPPLIER NAME, ISO)

ORDER LINE (ORDER, PRODUCT NUMBER)

CONVERTING 1NF TO 2NF

ORDER in 1NF

ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
86223	9-13-2015	633	Assembly	1	J-995	Jones	No
86223	9-13-2015	684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
86390	9-14-2015	304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

ORDER in 2NF

ORDER	ORDER DATE
86223	9-13-2015
86390	9-14-2015
86467	9-15-2015

PRODUCT in 2NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER	SUPPLIER NAME	ISO
128	Steel widget	A-602	Acme	Yes
304	Blue gadget	A-602	Acme	Yes
633	Assembly	J-995	Jones	No
684	Super gizmo	C-876	Cabot	Yes

ORDER LINE in 2NF

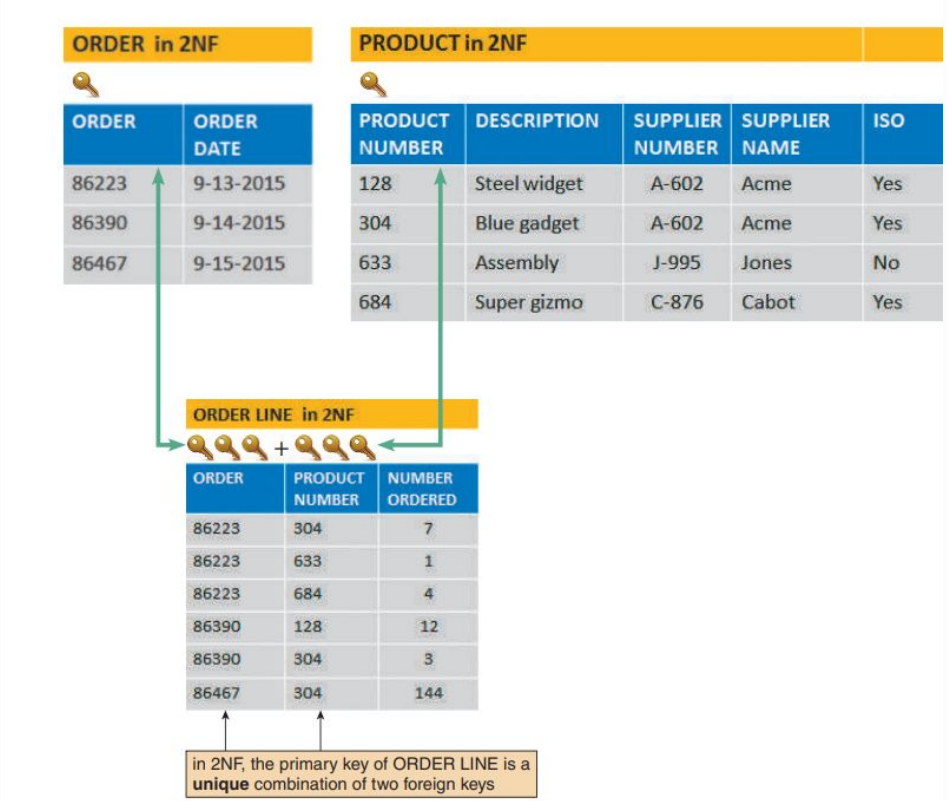
ORDER	PRODUCT NUMBER	NUMBER ORDERED
86223	304	7
86223	633	1
86223	684	4
86390	128	12
86390	304	3
86467	304	144

in 2NF, the primary key of ORDER LINE is a **unique** combination of two foreign keys

WHY IS IT IMPORTANT TO MOVE FROM 1NF TO 2NF?

Four kinds of problems are found with 1NF designs that do not exist in 2NF:

- Consider the work necessary to change a particular product's description.
- 1NF tables can contain inconsistent data.
- Adding a new product is a problem.
- Deleting a product is a problem.



THIRD NORMAL FORM

A design is in **Third Normal Form (3NF)** if every non-key field depends on..

- ***the key, the whole key, and nothing but the key.***

And if..

- it is in 2NF and
- no non-key field is dependent on another non-key field

3NF avoids redundancy and data integrity problems.

THIRD NORMAL FORM

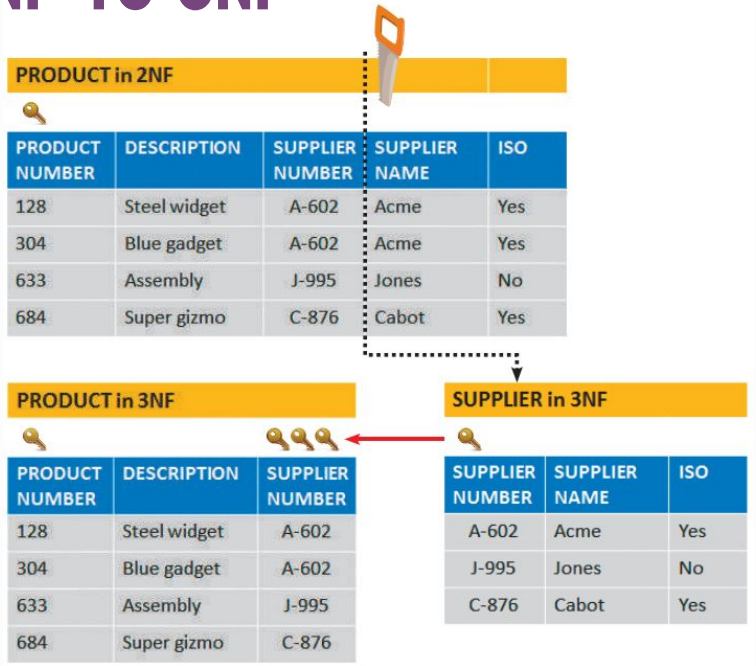
PRODUCT in 2NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER	SUPPLIER NAME	ISO
128	Steel widget	A-602	Acme	Yes
304	Blue gadget	A-602	Acme	Yes
633	Assembly	J-995	Jones	No
684	Super gizmo	C-876	Cabot	Yes

PRODUCT (PRODUCT NUMBER, DESCRIPTION, SUPPLIER NUMBER, SUPPLIER NAME, ISO)

CONVERTING 2NF TO 3NF

All fields from the 2NF table that depend on another non-key field must be removed and placed in a new table that uses the nonkey field as a primary key.



PRODUCT in 2NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER	SUPPLIER NAME	ISO
128	Steel widget	A-602	Acme	Yes
304	Blue gadget	A-602	Acme	Yes
633	Assembly	J-995	Jones	No
684	Super gizmo	C-876	Cabot	Yes

PRODUCT in 3NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER
128	Steel widget	A-602
304	Blue gadget	A-602
633	Assembly	J-995
684	Super gizmo	C-876

SUPPLIER in 3NF

SUPPLIER NUMBER	SUPPLIER NAME	ISO
A-602	Acme	Yes
J-995	Jones	No
C-876	Cabot	Yes

PRODUCT (PRODUCT NUMBER, DESCRIPTION, SUPPLIER NUMBER)

SUPPLIER (SUPPLIER NUMBER, SUPPLIER NAME, ISO)

DATA NORMALIZATION

ORDER (Unnormalized)

ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
		633	Assembly	1	J-995	Jones	No
		684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
		304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

ORDER in 2NF

ORDER	ORDER DATE
86223	9-13-2015
86390	9-14-2015
86467	9-15-2015

PRODUCT in 3NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER
128	Steel widget	A-602
304	Blue gadget	A-602
633	Assembly	J-995
684	Super gizmo	C-876

ORDER LINE in 2NF

ORDER	PRODUCT NUMBER	NUMBER ORDERED
86223	304	7
86223	633	1
86223	684	4
86390	128	12
86390	304	3
86467	304	144

SUPPLIER in 3NF

SUPPLIER NUMBER	SUPPLIER NAME	ISO
A-602	Acme	Yes
J-995	Jones	No
C-876	Cabot	Yes

07.

USING CODES

USING CODES

- A code is a set of letters or numbers that represents a data item. Codes can be used to simplify output, input, and data formats.
- Codes can represent data, they are encountered constantly in everyday life.

FOR EXAMPLE: STUDENT NUMBER

John Turner

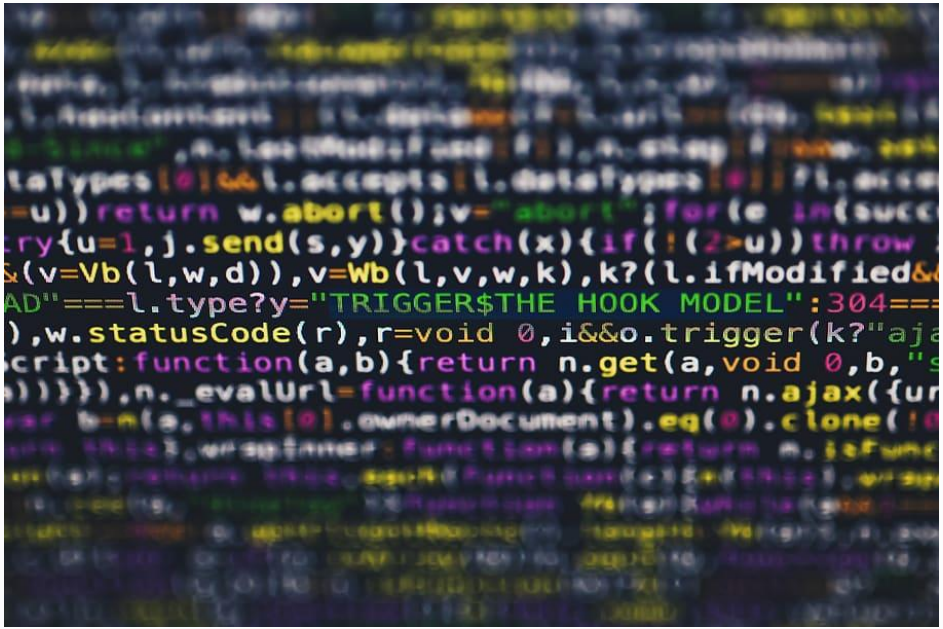
Student Number: 268960

John Turner

John Turner

USING CODES

- Codes can be used in many ways. Because codes are shorter than the data they represent, they save storage space and costs, reduce data transmission time, and decrease data entry time.
- Codes can reduce data input errors in situations when the coded data is easier to remember and enter than the original source data, when only certain valid codes are allowed, and when something within the code itself can provide immediate verification that the entry is correct.



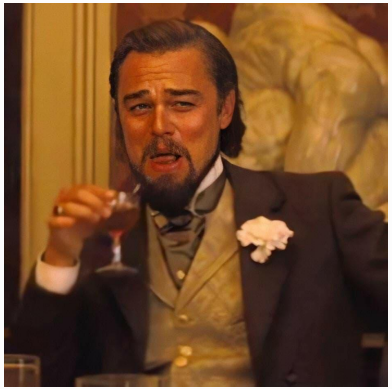
TYPES OF CODES

Sequence codes are numbers or letters assigned in a specific order. Sequence codes contain no additional information other than an indication of order of entry into the system.

For Example:

A human resource system issues consecutive employee numbers to identify employees.

EN: 584



EN: 433

TYPES OF CODES

Block sequence codes use blocks of numbers for different classifications.

For example:

College course numbers usually are assigned using a block sequence code.

100-level courses are freshman courses such as Chemistry 110, Mathematics 125.
200-level courses are sophomore courses such as History 220, Politics 250.

Additional meaning:

English 151 → English 152

TYPES OF CODES

Alphabetic codes use alphabet letters to distinguish one item from another based on a category, an abbreviation, or an easy-to-remember value, called a mnemonic code. Many classification codes fit more than one of the following definitions:

- **Category codes** identify a group of related items.

  For example: A local department store uses a two-character category code to identify the department in which a product is sold: GN for gardening supplies, HW for hardware, and EL for electronics.

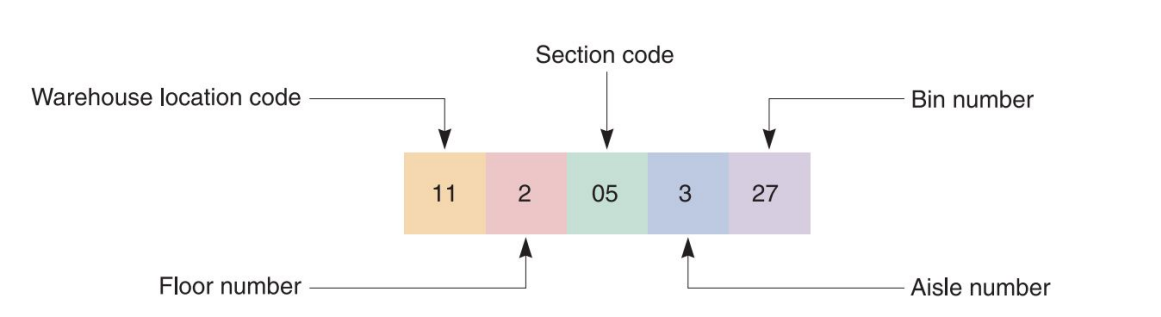
- **Abbreviation codes** are alphabetic abbreviations.

  For example: Standard state codes include NY for New York, ME for Maine, and MN for Minnesota. Some abbreviation codes are called mnemonic codes because they use a specific combination of letters that are easy to remember such as ATL for Atlanta and MIA for Miami.

TYPES OF CODES

Significant digit codes distinguish items by using a series of subgroups of digits

For example: Postal codes are significant digit codes.Other such codes include inventory location codes that consist of a two-digit warehouse code, followed by a one-digit floor number code, a two-digit section code, a one-digit aisle number, and a two-digit binnumber code. The inventory location code 11205327. What looks like a large eight-digit number is actually five separate numbers, each of which has significance



TYPES OF CODES

Derivation codes combine data from different item attributes, or characteristics. Most magazine subscription codes are derivation codes.

For example: One popular magazine uses a subscriber's five-digit postal code, followed by the first, third, and fourth letters of the subscriber's last name, the last two digits of the subscriber's house number, and the first, third, and fourth letters of the subscriber's street name. A sample is shown in Figure 9-36.

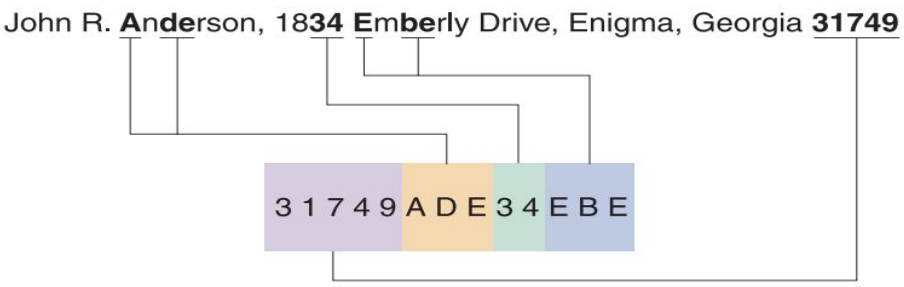


FIGURE 9-36 A magazine subscriber code is derived from various parts of the name and address.

TYPES OF CODES

Cipher codes use a keyword to encode a number.

For example: A retail store might use a 10-letter word, such as CAMPGROUND, to code wholesale prices, where the letter C represents 1, A represents 2, and so on. Thus, the code, GRAND, indicates that the store paid $562.90 for the item.

TYPES OF CODES

Action codes indicate what action is to be taken with an associated item.

For example: A student records program might prompt a user to enter or click an action code such as D (to display a record), A (to add a record), and X (to exit the program).

DESIGNING CODES

- **Keep codes concise**
  Do not create codes that are longer than necessary.
- **Allow for expansion**
  A coding scheme must allow for reasonable growth in the number of assigned codes.
- **Keep codes stable**
  Changes in codes can cause consistency problems and require data updates. During the changeover period, all the stored occurrences of a particular code and all documents containing the old code will have to change as users switch to the new code.
- **Make codes unique**
  Codes used for identification purposes must be unique to have meaning. If the code HW can indicate hardware or houseware, the code is not very useful.

- **Use sortable codes**
  If products with three-digit codes in the 100s or the 300s are of one type, while products with codes in the 200s are a different type, a simple sort will not group all the products of one type together
- **Use a simple structure**
  Do not code some part numbers with two letters, a hyphen, and one digit, and others with one letter, a hyphen, and two digits.
- **Avoid confusion**
  It is easy to confuse the number zero (0) and the uppercase letter O, or the number one (1) with the lowercase letter L (l) or uppercase letter I.

- **Make codes meaningful**
  Codes should be easy to remember, user-friendly, convenient, and easy to interpret.
- **Use a code for a single purpose**
  Do not use a code to classify unrelated attributes.

  For example, if a single code is used to identify the combination of an employee's department and the employee's insurance plan type, users will have difficulty identifying all the subscribers of a particular plan, or all the workers in a particular department, or both. A separate code for each separate characteristic makes much more sense

- **Keep codes consistent**
  For example, if the payroll system already is using two digit codes for departments, do not create a new, different coding scheme for the personnel system. If the two systems already are using different coding schemes, try to establish a consistent coding scheme.

08.

DATA STORAGE & ACCESS

Strategic Business Tools

TOOLS & TECHNIQUES

DATA WAREHOUSING

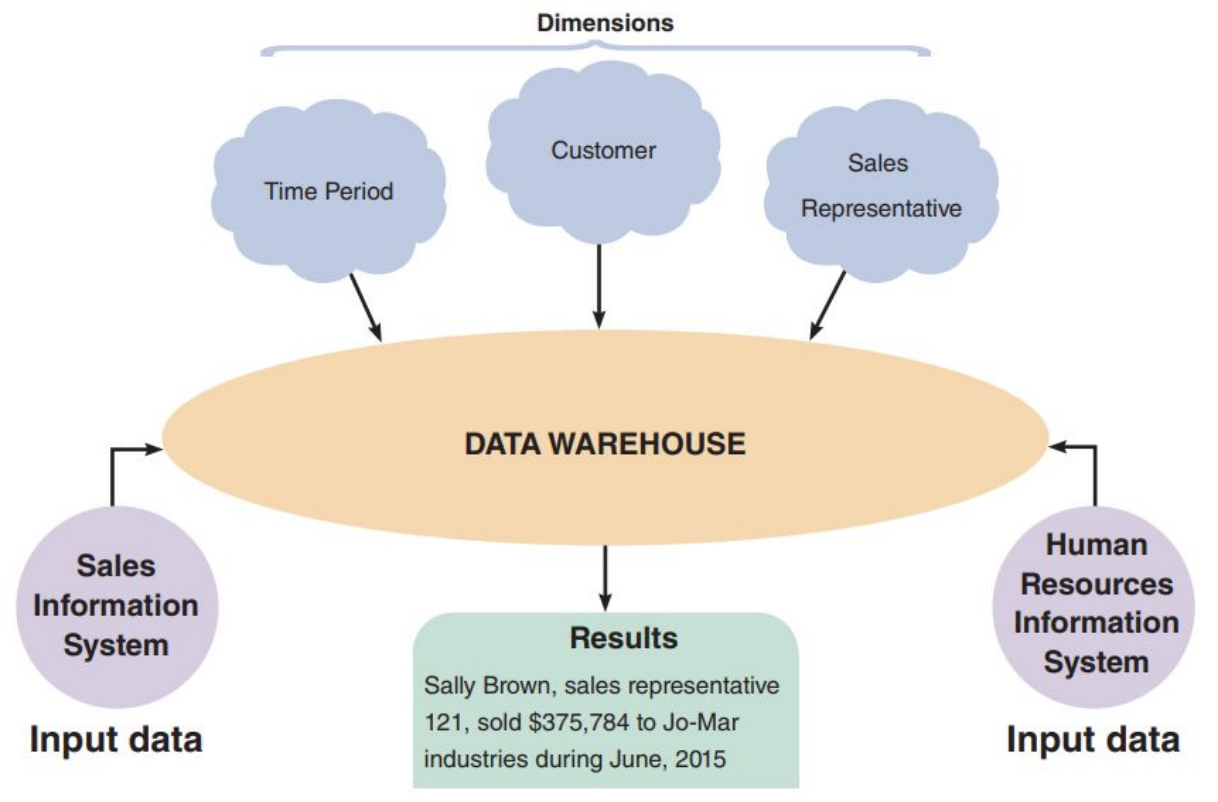
DATA MINING

DATA WAREHOUSING

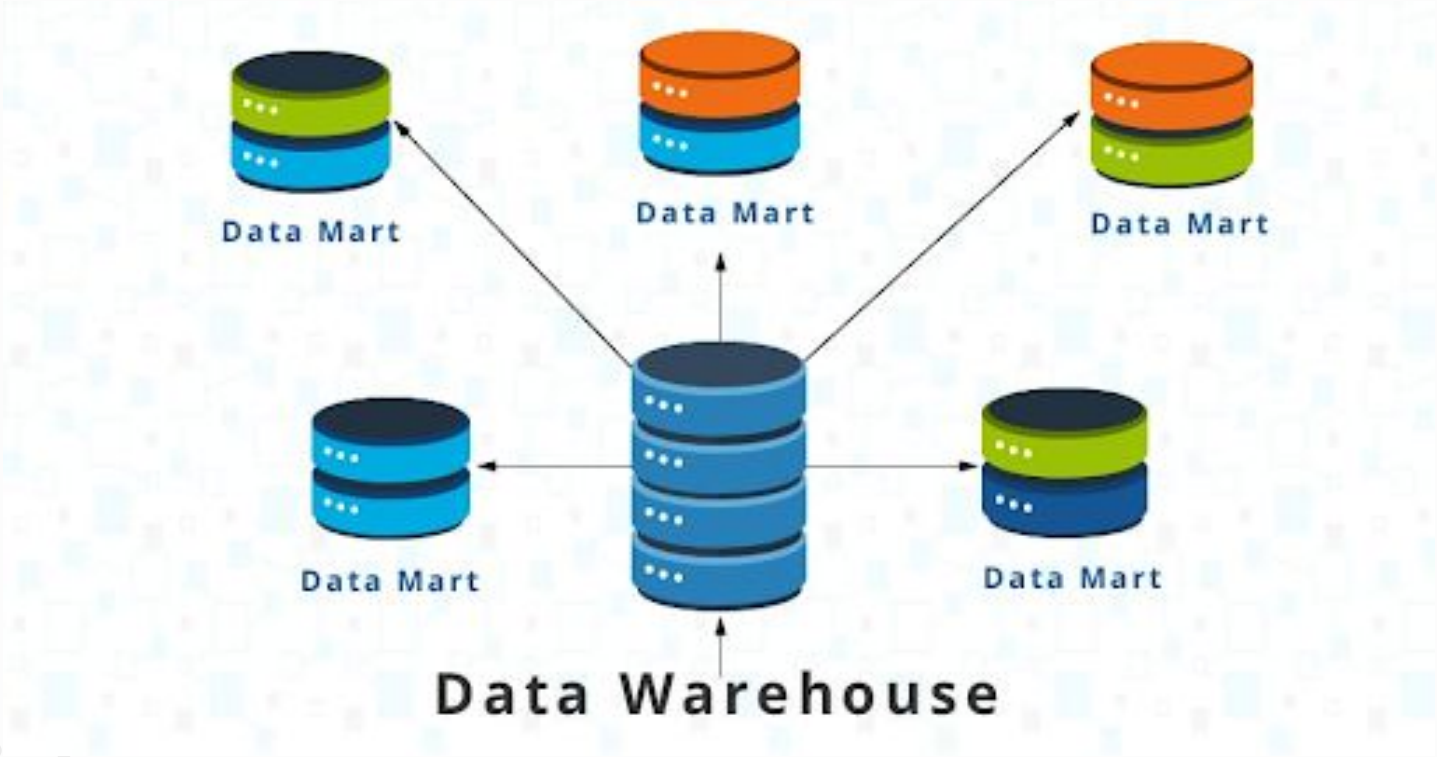
What is a data warehouse?

- Integrated collection of data
- Provides an enterprise-wide view to support management analysis and decision making
- Allows users to specify certain dimensions or characteristics

EXAMPLE: DATA WAREHOUSING



DATA MART


Data Mart
Data Mart
Data Mart
Data Mart
Data Mart
Data Warehouse

DATA MINING

What is data mining?

- Looking for meaningful data patterns and relationships
- Could help identify potential customers based on prior purchases
- Predict future trends and make more-informed business decisions
- “machine learning algorithms that find buried patterns in databases, and report or act on those findings.”
- However, this raises ethical and privacy issues

DATA MINING GOALS

- Increase the number of pages viewed per session
- Increase the number of referred customers
- Reduce clicks to close, which means average page views to accomplish a purchase or obtain desired information
- Increase checkouts per visit
- Increase average profit per checkout

MARKET BASKET ANALYSIS

- Understanding customer purchasing patterns
- Analyzing large data sets to reveal product groupings

STORAGE

LOGICAL

- Data a user can view, understand, and access
- Alphabetic and numeric characters
- Logical Record = set of field values that describes a single person, place, thing, or event

PHYSICAL

- Hardware-related
- Process of reading and writing binary data to physical media

DATA CODING

EBCDIC

Extended Binary Coded Decimal Interchange Code

ASCII

American Standard Code for Information Interchange

BINARY

Represents numbers as actual binary values

UNICODE

Coding standard that uses two bytes per character

STORING DATES

- Depends on how dates will be displayed
- Used only two digits at the beginning of the 21st century
- International Organization for Standardization (ISO) requires the format, YYYYMMDD
- Absolute date = total number of days from specific base date

09.

DATA CONTROL

Correct, Complete, and Secure Data Storage

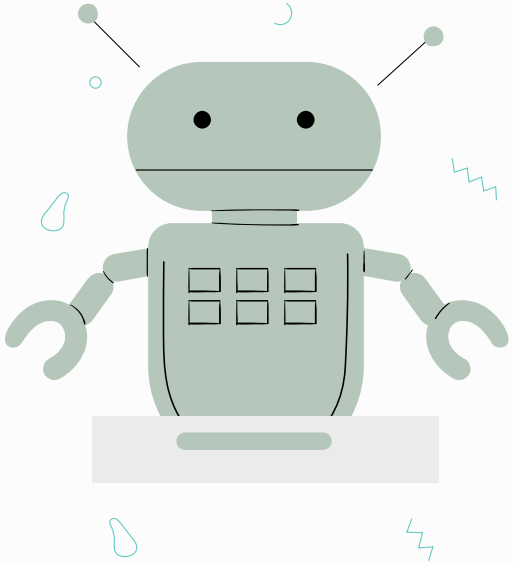
WELL-DESIGNED DBMS

BUILT-IN CONTROL & SECURITY FEATURES

Subschemas, passwords, encryption, audit trail files, backup and recovery procedures

ANALYST'S MAIN RESPONSIBILITY

Ensure that DBMS features are used properly



SECURITY FEATURES

SUBSCHEMA

Limited view to a specific user or level of users

PERMISSIONS

Different privileges

ENCRYPTION

Prevent unauthorized access

BACKUP

Backup and restore files

AUDIT LOG FILES

Record details of all accesses and changes

AUDIT FIELDS

Additional control or security information

10.

CHAPTER SUMMARY

CHAPTER SUMMARY

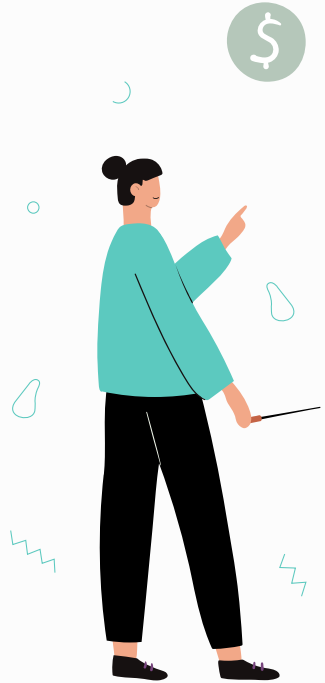
This chapter continued the study of the systems design phase of the SDLC. It was explained that files and tables contain data about people, places, things, or events that affect the information system.

- File-oriented systems, also called file processing systems, manage data stored in separate.
- A database consists of linked tables that form an overall data structure. A database management system (DBMS) is a collection of tools, features, and interfaces that enable users to add, update, manage, access, and analyze data in a database.
- Database environment offers scalability, support for organization-wide access, economy of scale, data sharing among user groups, balancing of conflicting user requirements, enforcement of standards, controlled redundancy, effective security, flexibility, better programmer productivity, and data independence.

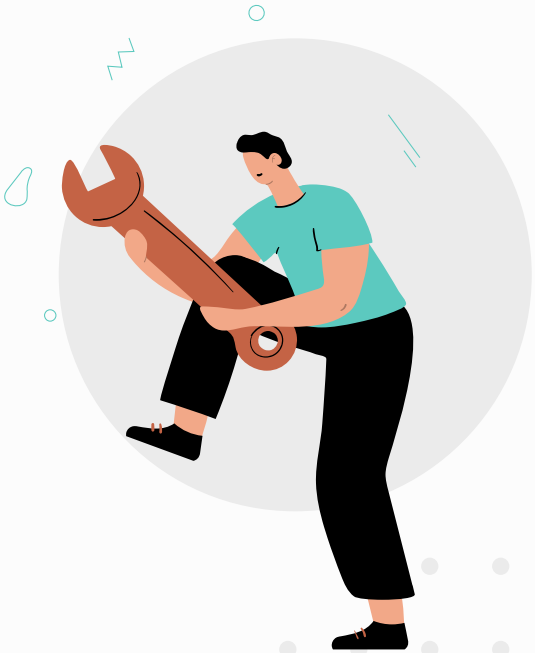
- A primary key is the field or field combination that uniquely and minimally identifies a specific record.
- A candidate key is any field that could serve as a primary key.
- A foreign key is a field or field combination that must match the primary key of another file or table.
- A secondary key is a field or field combination used as the basis for sorting or retrieving records.
- An entity-relationship diagram (ERD) is a graphic representation of all system entities and the relationships among them. The ERD is based on entities and data stores in DFDs prepared during the systems analysis phase.
- The three basic relationships represented in an ERD are one-to-one (1:1), one-to-many (1:M), and many-to-many (M:N).
- The relationship between two entities is also referred to as cardinality. A common form of cardinality notation is called crow's foot notation, which uses various symbols to describe the characteristics of the relationship.

- DBMS components include interfaces for users, database administrators, and related systems; a data manipulation language; a schema; and a physical data repository.
- Data management techniques include data warehousing, which stores data in an easily accessible form for user access, and data mining, which looks for meaningful patterns and relationships among data.
- Data mining also includes clickstream storage, which records how users interact with a site, and market basket analysis, which can identify product relationships and consumer buying patterns.
- Data design tasks include creating an initial ERD; assigning data elements to an entity; normalizing all table designs; and completing the data dictionary entries for files, records, and data elements.
- Code is a set of letters or numbers used to represent data in a system. Using codes can speed up data entry, reduce data storage space, and reduce transmission time.

    The main types of codes are sequence codes, block sequence codes, classification codes, alphabetic codes (including category codes, abbreviation codes, and mnemonic codes), significant digit codes, derivation codes, cipher codes, and action codes.



THANK YOU FOR LISTENING!



REFERENCE: SYSTEMS ANALYSIS AND DESIGN, IITH EDITION, CHAPTER 9