Shiksha Mandal's

G.S.College Of Commerce,Wardha

Department of B.Com Computer Application

BCCA Part-III Sem-V

System Analysis & Design

Presented By

Prof. Amol Raut
Assistant Professor

Dr. Revati Bangre
Co-Ordinator
BCCA Department
G.S.College of Commerce,Wardha

Dr. Anil Ramteke
Principal
G.S.College of Commerce, Wardha

UNIT- III

System Design- The Process And Stages Of System Design- Introduction, The Process Of Design, Design Methodologies, Major Development Activities, Audit Consideration.

Input/output And Forms Design- Introduction, Input Design, Output Design, Forms Design.

File Organization And Data Base Design- Introduction, File Structure, File Organization, Data Base Design, The Role Of The Data Base Administrator.

System Design- The Process And Stages Of System Design

The Process of Design

In the system development life cycle. User requirements have been identified. Information has been gathered to verify the problem and evaluate the existing system. A feasibility study has been conducted to review alternative solutions and provide cost/benefit justification. The culmination of the study is a proposal summarizing the finding and recommending a candidate system for the user.

At this point in the systems life cycle, the design phase begins. The design is a solution a how to approach compared to analysis a what is orientation. It translates the system requirements into ways of operationalizing them. It covers the process and stages of systems design, the tools used to design candidate systems and the user input to system design.

- The design phase focuses on the detailed implementation of the system recommended in the feasibility study. Emphasis is on translating performance specifications into design specifications.
- The design phase is a transition from a user-oriented document to a document oriented to the programmers or data base personnel.

Logical and Physical Design

❖ Systems design goes through two phases of development: logical and physical design. A data flow diagram shows the logical flow of a system and defines the boundaries of the system. For a candidate system it describes the inputs, outputs, data bases and procedures.

❖ All in a format that meets the user's requirements. When analysts prepare the logical system design, they specify the user needs at a level of detail that virtually determines the information flow into and out of the system and required data resources.

The design covers the following:

1.Reviews the current physical system-its data flows, file content, volumes, frequencies.

2.Prepares output specifications-that is, determines the format, content, and frequency of reports, including terminal specifications and locations.

3.Prepares input specification- format, content and most of the input functions. This includes determining the flow of the document from the input data source to the actual input location.

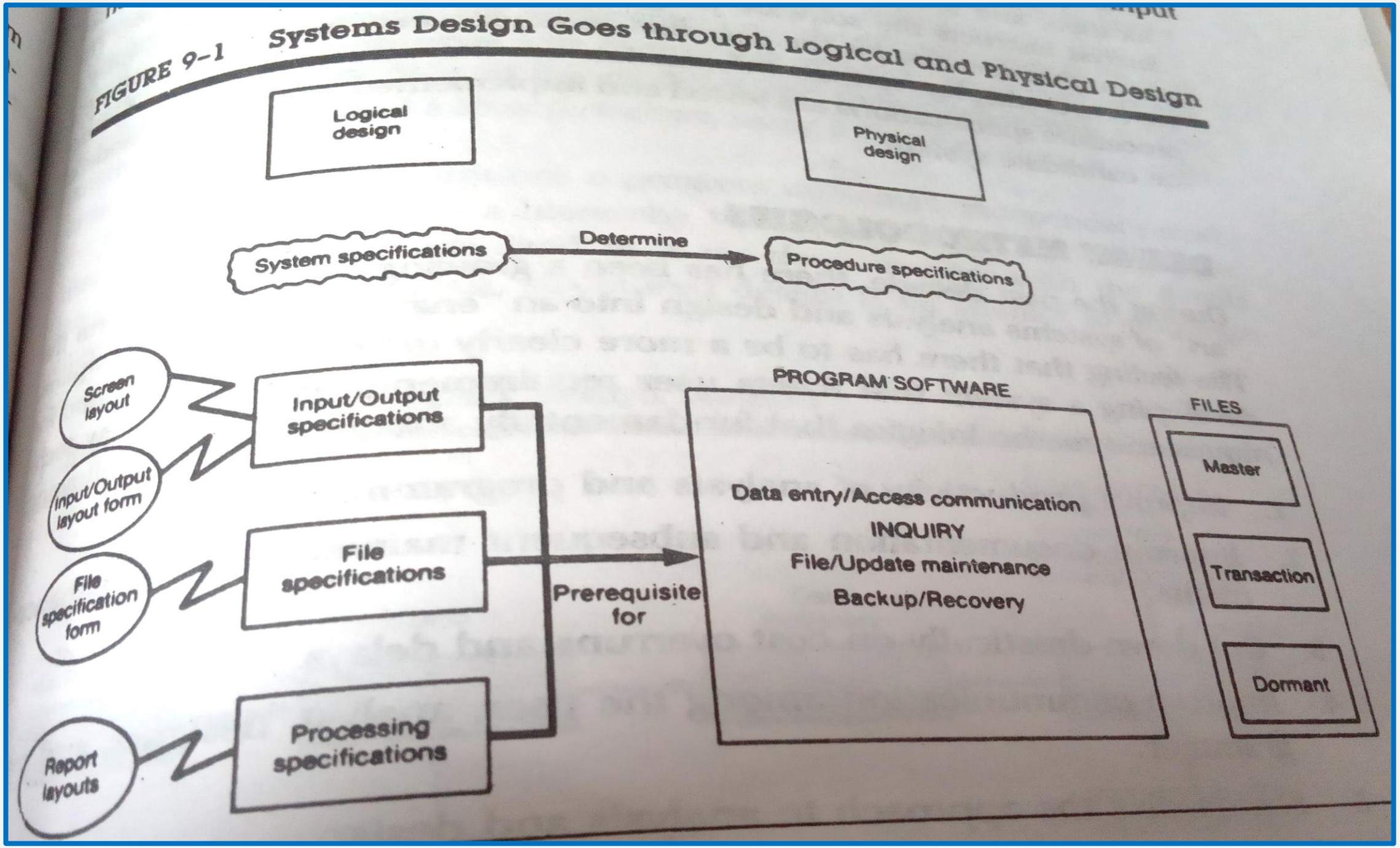
4.Prepares edit, security and control specifications. This includes specifying the rules for edit correction, backup procedures, and the controls that ensure processing and file integrity.

5.Specifies the implementation plan.

6.Prepares a logical design walkthrough of the information flow, output, input, controls, and implementation plan.

7.Reviews benefits, costs, target dates and system constraints.

FIGURE 9–1 Systems Design Goes through Logical and Physical Design



The logical design also specifies output, input, file and screen layouts. In contrast, procedure specifications show how data are entered, how files are accessed, and how files are accessed, and how files are accessed, and how reports are produced.

Physical system design consists of the following steps:

1.Design the physical system

a. Specify input/output media.

b. Design the data base and specify backup procedures.

c. Design physical information flow through the system and a physical design walkthrough.

2.Plan system implementation

a. Prepare a conversion schedule and a target date.

b. Determine training procedure, courses and timetable.

3.Devise a test and implementation plan and specify any new hardware/software.

4.Update benefits, costs, conversion date and system contraints.

Design Methodologies

There has been a growing move to transform the art of systems analysis and design into an engineering type discipline. The feeling that there has to be a more clearly defined logical method for developing a system that meets user requirements has led to new techniques and methodologies that fundamentally attempt to do the following.

1.Improve productivity of analysts and programmers.

2.Improve documentation and subsequent maintenance and enhancements.

3.Cut down drastically on cost overruns and delays.

4.Improve communication among the user, analyst, designer and programmer.

5.Standardize the approach to analysis and design.

6.Simplify design by segmentation.

Structured Design

- Structured design is a data-flow based methodology. The approach begins with a system specification that identifies inputs and outputs and describes the functional aspects of the system.
- The system specification, then are used as a basis for the graphic representation- data flow diagram (DFD) of the data flows and processes.
- From the DFD, then next step is the definition of the modules and their relationships to one another in a form called a structure chart, using a data dictionary and other structured tools.
- Structured design partitions a program into small, independent modules. They are arranged in hierarchy that approximates a model of the area and is organized in a top-down manner with the details shown at the bottom. Thus, structured design is an attempt to minimize complexity and make a problem manageable by subdividing it into smaller segments, which is called modularization or decomposition this way, structuring minimizes intuitive reasoning and promotes maintainable provable systems.

A design is said to be top-down if it consists of a hierarchy of modules, with each module having a single entry and single exit subroutine. The primary advantages of this design are as follows:

1.Critical interfaces are tested first.

2.Early versions of the design, through incomplete are useful enough to resemble the real system.

3.Structuring the design provides control and improves morale.

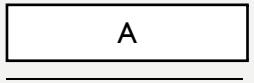
4.The procedural characteristics define the order that determines processing.

So structured design arises from the hierarchical view of the application rather than the procedural view. The top level shows the most important division of work, the lowest level at the bottom shows the details.

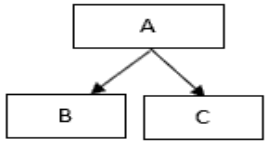
Functional Decomposition

The documentation tool for structured design is the hierarchy or structure chart. It is a graphic tool for representing hierarchy, and it has three elements.

1.The module is represented by a rectangle with a name. It is a contiguous set of statements.

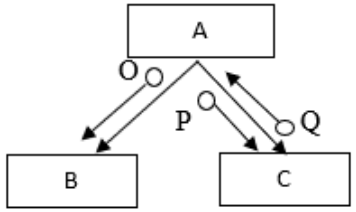


2, The connection is represented by a vector linking two modules. It usually means one module has called another module. Module A calls module B, it also calls module C.



3.The couple is represented by an arrow with a circular tail. It represents data items moved from one module to another. O, P and Q are couples. Module A calls B, passing O downward. Likewise,

Module A calls C passing P downward and receiving Q back.



Major Development Activities

Several development activities are carried out during structured design. They are data base design, implementation planning, system test preparation system interface specification and user documentation.

1.Database design: This activity deals with the design of the physical database. A key is to determine how the access paths are to be implemented. A physical path is derived from a logical path. It may be implemented by pointers chains or other mechanisms.

2.Program design: In conjunction with database is a decision on the programming language to be used and the flowcharting, coding and debugging procedure prior to conversion. The operating system limits the programming languages that will run on the system. When the system design is done, the plans and test cases for implementation are required. So, there must be detailed schedules for system testing and training of user staff.

3.System and Program test preparation: Each aspect of the system has a separate test requirement. System testing is done after all programming and testing are completed. The test cases cover every aspect of the candidate system- Acceptance testing is another testing that convinces the user that the candidate system will meet the stated requirements. It is conducted in the presence of users, audit representatives or the entire staff.

4.System interface specification: This phase specifies for the user the way in which information should enter and leave the system. The designer offers the user various options.

Before the system is ready for implementation, user documentation in the form of a user or operator's manual must be prepared. The manual provides instructions on how to access, update, or retrieve information, how to display or print output, in what format and so on.

Personnel Allocation:

The structured approach to design and implementation is useful in facilitating the planning process. Emphasis is on allocating the right programmers to the task within a realistic timetable. A completed structure chart gives the designer a realistic outline of an idea of the work to be done. Programmers are the assigned appropriately. Programmers are assigned subsystems that are strongly cohesive and loosely coupled. Once modules are allocated, roles are allocated within each team and the designer oversees all the work. Assigning modules are very important. Modules at the bottom are important because they represent the user interface. So, a team with specialized skills should be assigned to such a module.

Audit Considerations

Audit considerations examine the results of the analysis by using both the narratives and models to identify the problems caused due to misplaced functions, split processes or functions, broken data flows, missing data, redundant or incomplete processing, and no addressed automation opportunities.

The activities under this phase are as follows −

- Identification of the current environment problems
- Identification of problem causes
- Identification of alternative solutions
- Evaluation and feasibility analysis of each solution
- Selection and recommendation of most practical and appropriate solution
- Project cost estimation and cost benefit analysis

Input Design

- Inaccurate input data are the most common cause of errors in data processing. Errors entered by data entry operators can be controlled by input design. Input design is the process of converting user-originated inputs to a computer-based format. Input In an information system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.

- Therefore, the quality of system input determines the quality of system output. Well-designed input forms and screens have following properties −

  - It should serve specific purpose effectively such as storing, recording, and retrieving the information.
  - It ensures proper completion with accuracy.
  - It should be easy to fill and straightforward.
  - It should focus on user's attention, consistency, and simplicity.
  - All these objectives are obtained using the knowledge of basic design principles regarding
    - What are the inputs needed for the system?
    - How end users respond to different elements of forms and screens.

Input Data

The goal of designing input data is to make data entry as easy, logical, and free from errors as possible. In entering data, operators need to know the following:

1.The allocated space for each field.

2.Field sequence, which must match that in the source document.

3.The format in which data fields are entered, for example, filling out the date field is required through the edited format mm/dd/yy.

Objectives for Input Design

The objectives of input design are −

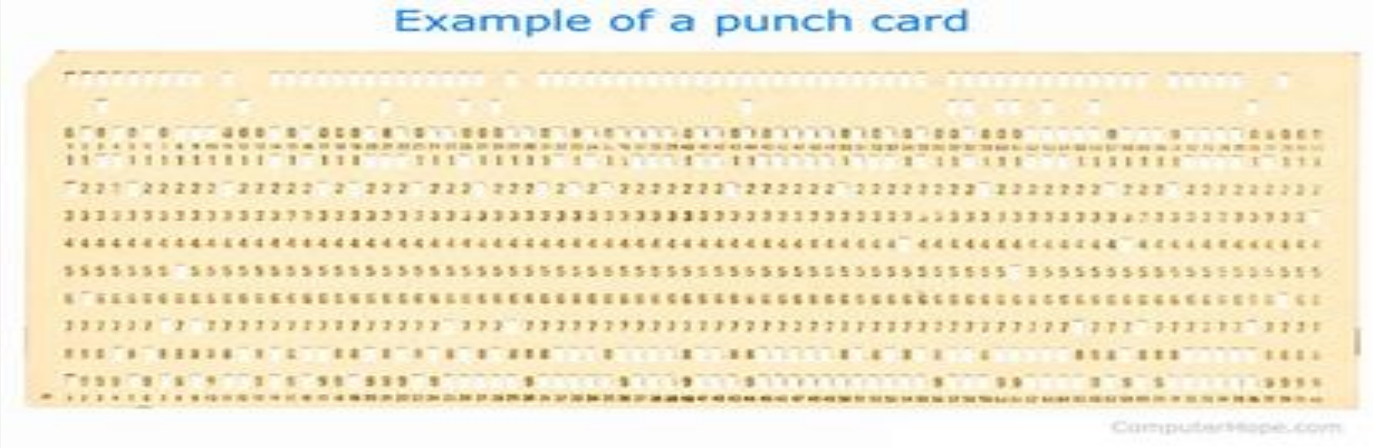
1. To design data entry and input procedures
2. To reduce input volume
3. To design source documents for data capture or devise other data capture methods
4. To design input data records, data entry screens, user interface screens, etc.
5. To use validation checks and develop effective input controls.

Input media and devices

a device that can be used to insert data into a computer or other computational device.

1.A punched card or punch card :is a piece of stiff paper that can be used to contain digital data represented by the presence or absence of holes in predefined positions. Digital data can be used for data processing applications or, in earlier examples, used to directly control automated machinery.

Punched cards were widely used through much of the 20th century in the data processing industry, where specialized and increasingly complex unit record machines, organized into semiautomatic data processing systems, used punched cards for data input, output, and storage. Many early digital computers used punched cards, often prepared using keypunch machines, as the primary medium for input of both computer programs and data.



2)Key-to diskette: Alternatively referred to as a floppy or floppy disk, a floppy diskette is a type of storage media, capable of storing electronic data, like a computer file. The floppy diskette was first created in 1967 by IBM as an alternative to buying hard drives, which were extremely expensive at the time.

The picture shown on this page is an example of a 3.5" floppy diskette, which was one of the most commonly used floppy diskettes, capable of storing 1.44 MB of data.

MICR: The Magnetic Ink Character Recognition (MICR) system reads characters printed in a special magnetic ink into the computer. The main users of MICR are banks. They use it to read information from cheques into their computers so that the cheques can be cashed. Here is some information stored on a cheque using MICR.

The information printed on the cheque using MICR is :

- A unique number for the cheque.
- A code that identifies the bank and branch that issued the cheque (the sort code).
- The number of the account that the cheque relates to.

MICR readers can only read one special font which can represent only numbers and a few punctuation marks. They can read characters very quickly and with 100% accuracy. Information printed in magnetic ink is also very secure. It is not possible to change the information by writing over it with a pen and the printed numbers are not damaged by folding (as often happens with cheques). Both the reader used by MICR and the special ink are expensive.

3)Mark sensing readers: A method for data input in which electrically conductive marks, usually made with a soft graphite pencil on a preformatted card or form, are electrically sensed. This method has been displaced by the more reliable method of OMR (i.e. optical mark reading) in which the marks are detected photo electrically.

4)Optical bar code reader:

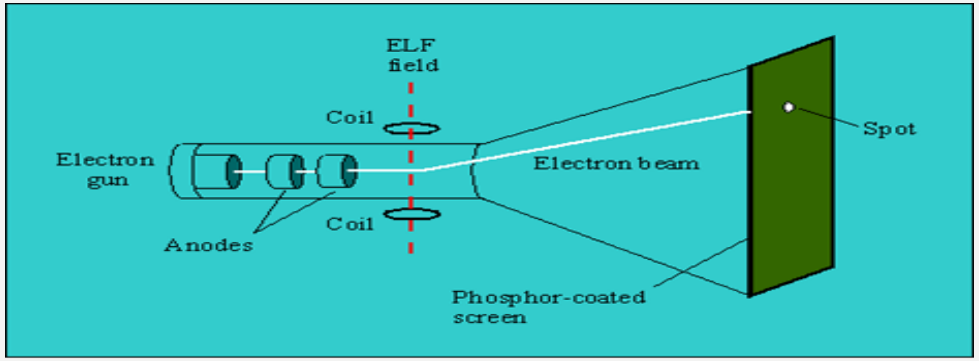
A barcode reader is a hand-held or stationary input device used to capture and read information contained in a barcode. A bar-code reader consists of a a lens, light source and a light sensor which translates optical impulses into electrical ones. Moreover, nearly all barcode readers consists of a decoder circuitry that analyzes the barcode’s image data provided by the sensor and sends the barcode’s content to the scanner’s output port.

5)Cathode ray tube (CRT):

A cathode ray tube (CRT) is a specialized vacuum tube in which images are produced when an electron beam strikes a phosphorescent surface. Most desktop computer displays make use of CRTs. The CRT in a computer display is similar to the “picture tube" in a television receiver.

A cathode ray tube consists of several basic components, as illustrated below. The electron gun generates anarrow beam of electrons.

The anodes accelerate the electrons. Deflecting coils produce an extremely low frequency electromagnetic field that allows for constant adjustment of the direction of the electron beam. There are two sets of deflecting coils: horizontal and vertical. (In the illustration, only one set of coils is shown for simplicity.) The intensity of the beam can be varied. The electron beam produces a tiny, bright visible spot when it strikes the phosphor-coated screen.



Online data entry:

We live in the age of microcomputer and at a time where more and more CRT screens are used for online data entry. Now terminal(computers) prices are falling and with the help of keyboard it is very easy to to input. the no. of applications used have a long list.

Such as ATMs, online data entry makes use of processors that accepts commands and data from the operator through keyboard devices such as touch sensitive screens, or voice input. The input received by the processor and accepted and rejected.

There are three major approaches for entering the data into computer

a) Menu
b) Formatted form
c) Prompts

a) Menu:

A menu is a selection of list which simplifies computer data access or entry. Instead of remembering what to enter from a list of options and types the option letter associated with it. for example the below list shows a menu for entering the city by options.

- Nagpur
- Wardha
- Amravati

b) Formatted form:

A formatted form is preprinted form or a template that request the user to enter data in appropriate location. it is fill in the blanks type form. The form is flashed on screen as a screen as a unit. The cursor is usually positioned at the first blank, after the user responds the cursor automatically moves to the next line and so on until the form is completed during user may move cursor anywhere up, down, left and right.

Saving form

a/c no.:

name of a/c holder:

rs……..

in words…….

Sign:

c)The prompt: At prompt the system displays one inquiry at time asking the user for a response. for ex.the following dialogue represents a prompt approach to data entry. z

System: enter ID

User: Password

Most systems edit data entered by the user. for example, if the password exceeds a maximum no .of digits or the password is illegal he system responds with the message like" unauthorized entry". the user has three chances after the system locks. The prompt method allows the user the key questions that determines the next response of time.

Output Design

The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts.

Objectives of Output Design

The objectives of input design are −

- To develop output design that serves the intended purpose and eliminates the production of unwanted output.
- To develop the output design that meets the end users requirements.
- To deliver the appropriate quantity of output.
- To form the output in appropriate format and direct it to the right person.
- To make the output available on time for making good decisions.

Form design:

People read from forms, write on forms and spend hours handling forms and filing forms. The data the forms carry come from people and the informational output goes to people. Form is a tool with a message. It is the physical carrier of data-of information. It is an either an authority for action or a request for action.

Classification of forms:

A printed form is generally classified by what it does in the system. There are three primary classifications

1.Action: This type of form requests the user to do something.

Example: purchase orders.

2.Memory: This form is a record of historical data that remains in a file, is used for reference, and serves as control on key details.

Example: Inventory records, purchase records

3.Report: This form guides supervisors and other administrators in the activities. It provides data on a project or a job.

Example: profit and loss statements, sales analysis report

Requirements of form design:

Form design follows analyzing forms. Since the purpose of a form is to communicate effectively through forms design, there are several major requirements.

1.Identification and wording: The form title must clearly identify its purpose. Columns and rows should be labeled to avoid confusion. The form should also be identified by firm name or code number to make it easy to reorder.

2.Maximum readability and use: The form must be easy to use and fill out. It should be legible, intelligible and uncomplicated. Ample writing space must be provided for inserting data.

3.Physical factors: The forms composition, color, layout and paper stock should lend themselves to easy reading. Pages should be numbered when multipage reports are being generated for the user.

4.Order of data items: The data requested should reflect a logical sequence. Related data should be in adjacent positions. Data copied from source documents should be in the same sequence on both forms.

5.Ease of data entry: If used for data entry, the form should have field positions indicated under each column of data and should have some indication of where decimal points are.

6.Size and arrangement: The form must be easily stored and filed. It should provide for signatures. Important items must be in a prominent action on the form.

7.Use of instructions: The instructions that accompany a form should clearly show how it is used and handled.

8.Efficiency considerations: The form must be cost effective. This means eliminating unnecessary data and facilitating reading lines across the form.

9.Type of report: Forms design should also consider whether the content is executive summary, intermediate managerial information, or supporting data. The user requirements for each type determine the final form design.

Carbon paper as form copier:

Carbon paper is one way of duplicating information in a form. There are two types of carbon, classified by the action they encounter

Various methods of transferring impressions between copies are as follows

1.One-time carbon: It is made of inexpensive Kraftex paper. It is interleaved between two sheets in the form. It is used once and then thrown away. It is the most cost-effective for multipart forms.

2.Carbon backed paper: The back of each form copy is coated with carbon, which transfers data to the copy beneath.

3.NCR (No Carbon Required) paper: The top sheet is chemically treated with invisible dye, that allows impressions to be transferred to the next lower copy. It is the cleanest and the costliest method. Erasing removes the coating permanently.

The readability of the carbon copies depends on the color and outline. In multiple copies, the copies below the original should be lighter in weight for easy transfer of the carbon. Generally, one time carbon is preferred when a small number of copies are required. If carbon is unacceptable, NCR can be used.

Objectives of Good Form Design

- A good form design is necessary to ensure the following −
- To keep the screen simple by giving proper sequence, information, and clear captions.
- To meet the intended purpose by using appropriate forms.
- To ensure the completion of form with accuracy.
- To keep the forms attractive by using icons, inverse video, or blinking cursors etc.
- To facilitate navigation.

Types of Forms

Flat Forms

- It is a single copy form prepared manually or by a machine and printed on a paper. For additional copies of the original, carbon papers are inserted between copies.
- It is a simplest and inexpensive form to design, print, and reproduce, which uses less volume.

Unit Set/Snap out Forms

- These are papers with one-time carbons interleaved into unit sets for either handwritten or machine use.
- Carbons may be either blue or black, standard grade medium intensity. Generally, blue carbons are best for handwritten forms while black carbons are best for machine use.

Continuous strip/Fanfold Forms

- These are multiple unit forms joined in a continuous strip with perforations between each pair of forms.
- It is a less expensive method for large volume use.

No Carbon Required (NCR) Paper

- They use carbonless papers which have two chemical coatings (capsules), one on the face and the other on the back of a sheet of paper.
- When pressure is applied, the two capsules interact and create an image.

Layout considerations in forms:

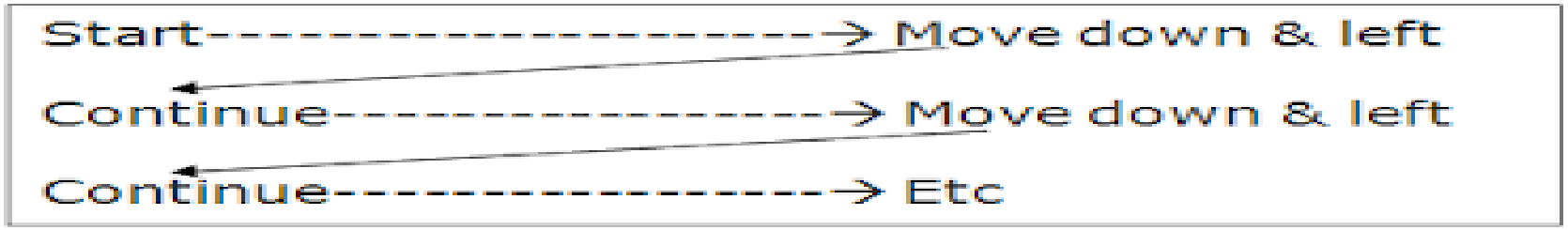
When a form is designed, a list is prepared of all the items to be included on the form and the maximum space to be reserved. The list should be checked by the form to make sure it has the required details.

1.Form title and number: The first consideration in forms design is a brief, descriptive title that tells what the form is and what it does. Since we read from left to right and from top to bottom, the upper left corner of the form is an appropriate place or a title. On forms that go outside the organization, the title is placed in the center at the top of the form. Long titles with vague words should be avoided. Good titles often are no longer than two words.

2.**Data classification and zoning**: All items in the form should be logically grouped. Then data is placed in the appropriate zones. Thus
A form is divided into zones (zone represents a group of similar information).

The zones use ULC (Upper Left Corner) method.

When more than one form is involved, the sequence of data in related forms should follow the same flow.



3.Rules and captions: In designing forms, use rules (lines) to guide the human eye to read and write data groups. A caption is similar to a column heading. It specifies what information to write in the space provided. Rules and captions go together. Rules guide and separate where as captions guide and instruct. Since a caption is used to guide, one or two different sizes of captions are usually used. A caption should not be very bold. Light rules should be used unless there is a definite need for separating parts of the form. In that case, heavier printed rules can be used. A column heading is a caption used to refer to more than one rule or box on a form.

A form is designed with a combination of rules and captions. Rules can also be used to make boxes in which the user places data. The caption tells the user what information goes in a particular position.

4.Box design: Whenever possible, it is advisable to design the form using the box style rule, with captions in the upper left corner. The box design gets the captions up out of the way and reduces the form size by 25 to 40 percent. It also makes the data entry uninterrupted from left to right.

5.Spacing Requirements: In a form there must be sufficient space to allow for data capture. A most commonly used standard is called 3/5 spacing. The 3 refers to the number of lines per vertical inch, while the 5 refers to the number of characters that fit in one horizontal inch. There are times when the number of lines have to be a minimum. This is done using the diagonal spacing method.

6.Ballot box and check off designs: Using ballot or check off boxes for questions that can be answered by yes or no can greatly reduce the amount of required writing. The user indicates the preference simply by checking off the desired box or placing a mark in the appropriate space.

7.Form instructions: A well-designed form with clearly stated captions should be self-instructing. A form becomes self-instructing by means of clearly stated captions and brief, procedural instructions.

8.Paper selection: Forms may be printed on paper of different colors, grades and weights. Colored paper or colored printing on white paper is used to distinguish among copies and facilitate sorting copies. Common color preferences are listed in the table.

Order of copy	color
First	white
Second	yellow
Third	pink
Fourth	blue

Paper weight is also to be considered. There are three major factors in paper selection

1.Appearance
2.Longevity
3.Handling.

- The form designer should know the number of times the form will be handled, the amount of folding it will receive and the extent to which it will be exposed.
- **Paper is classified as follows**

- Onion skin paper is used for inner copies of multiple part sets.
- Bond paper is rag paper that is the best.
- Duplicator paper is used for duplicating and Xerox machines.
- Ledger paper is used for checks, accounting records and ledger cards.
- Index paper is for printing cards.
- Card stock is the heaviest and the lowest grade paper.

File Structure

The terms cover is byte, data item, record, file and database. A byte is an arbitrary set of eight bits that represent a character. It is the smallest addressable unit in today's computers.

- **Data item:** one or more bytes are combined into a data item to describe an attribute of an object. For example, if the object is an employee. A data item is sometimes referred to as a field.

- **Record:**

The data items related to an object are combined into a record.

File:

- A collection of related records makes up a file. The size of a file is limited by the size of memory or the storage medium. Two characteristics determine how files are: activity and volatility.

Database:

- The highest level in the hierarchy is the database. It is a set of interrelated files for real-time processing. It contains the necessary data for problem solving and can be used by several users accessing data concurrently.

FILE ORGANISATION:

- **File** organization" refers to the logical relationships among the various records that constitute the **file**, particularly with respect to the means of identification and access to any specific record. "**File structure**" refers to the **format** of the label and data blocks and of any logical record control information refers to the way data is stored in a **file**. **File organization** is very important because it determines the **methods** of access, efficiency, flexibility and storage devices to use. There are four **methods** of **organizing** files on a storage media.
- File organization refers to the way data is stored in a file. File organization is very important because it determines the methods of access, efficiency, flexibility and storage devices to use. There are four methods of organizing files on a storage media.
- A file is organized to ensure that records are available for processing.it should be designed in line with the activity and volatility of the information and nature of the storage media and devices.

1.Cost of the media

2.Inquery requirements

3.File privacy, integrity, security and confidentiality.

There are four methods of organizing files: sequential, indexed-sequential, inverted list and direct access. File organization refers to the way data is stored in a file. File organization is very important because it determines the methods of access, efficiency, flexibility and storage devices to use. There are four methods of organizing files on a storage media. This includes:

- Sequential
- Random or Direct access file Organization
- Serial and
- indexed-sequential

1. Sequential file organization

- Sequential organization simply means storing and sorting in physical, contiguous blocks within files on tape or disk.
- Records are also in sequence within each block. To access a record, previous records within the blocks are scanned. Thus, sequential record design is best suited for “get next” activities, reading one record after another without a search delay.
- In a sequential organization, records can be added only at the end of the file. It is not possible to insert a record in the middle of the file without rewriting the file.
- In a data base system, however, a record may be inserted anywhere in the file, which would automatically resequence the records following the inserted record.
- Records are stored and accessed in a particular order sorted using a key field.

- Retrieval requires searching sequentially through the entire file record by record to the end.
- Because the record in a file are sorted in a particular order, better file searching methods like the binary search technique can be used to reduce the time used for searching a file.
- Since the records are sorted, it is possible to know in which half of the file a particular record being searched is located, Hence this method repeatedly divides the set of records in the file into two halves and searches only the half on which the records is found.
- For example, of the file has records with key fields 20, 30, 40, 50, 60 and the computer is searching for a record with key field 50, it starts at 40 upwards in its search, ignoring the first half of the set.

Advantages of sequential file organization

- The sorting makes it easy to access records.
- The binary chop technique can be used to reduce record search time by as much as half the time taken.
- Simple to design.
- Easy to program.
- Variable length and blocked records available.
- Best use of storage space.

Disadvantages of sequential file organization

- The sorting does not remove the need to access other records as the search looks for particular records.
- Sequential records cannot support modern technologies that require fast access to stored records.
- The requirement that all records be of the same size is sometimes difficult to enforce.
- Records cannot be added at the middle of the file.

Random or Direct access file organization

- In direct-access file organization, records are placed randomly throughout the file. Records need not be in sequence because they are updated directly and rewritten back in the same location.
- New records are added at the end of the file or inserted in specific locations based on software commands.
- Records are accessed by addresses that specify their disk locations. An address is required for locating a record, for linking records, or for establishing relationships.
- Records are accessed by addresses that specify their disk locations. An address is required for locating a record, for linking records, or for establishing relationships. Addresses are of two types:
- **Absolute**
- **Relative.**

- **A absolute address** represents the physical location of the record. It is usually stated in the format of sector/track/record number. One problem with absolute address is that they become in valid when the file that contains the records is relocated on the disk.
- **A relative address** gives a record location relative to the beginning of the file. There must be fixed length records for reference. Another way of locating a record is by the number of bytes it is from the beginning of the file.
- When the file is moved, pointers need not be updated because the relative location remains the same.
- Records are stored randomly but accessed directly.
- To access a file stored randomly, a record key is used to determine where a record is stored on the storage media.
- **Magnetic** and **optical** disks allow data to be stored and accessed randomly.

Advantages of random file access

- Quick retrieval of records.
- The records can be of different sizes.
- Records can be inserted or updated in the middle of the file.
- Better control over record allocation.

Disadvantages of Direct file organization

- Calculating address required for processing.
- Impossible to process variable length records.

Inverted list or Serial file organization

- Like the indexed- sequential storage method the inverted list organization maintains an index. The two methods differ, however, in the index level and record storage.
- The indexed sequential method has a multiple index for a given key, where as the inverted list method has a single index for each key type. In an inverted list, records are not necessarily stored in a particular sequence.
- They are placed in the data storage area, but indexes are updated for the record key and location.
- The inverted keys are best for applications that request specific data on multiple keys. They are ideal for static files because additions and deletions cause expensive pointer updating.
- Records in a file are stored and accessed one after another.
- The records are not stored in any way on the storage medium this type of organization is mainly used on **magnetic tapes**.

Advantages of serial file organization

- It is simple.
- It is cheap.
- Used in applications requesting specific data on multiple keys.

Disadvantages of serial file organization

- It is cumbersome to access because you have to access all proceeding records before retrieving the one being searched.
- Wastage of space on medium in form of inter-record gap.
- It cannot support modern high-speed requirements for quick record access.

Indexed-sequential file organization method

- Sequential organization, keyed sequential organization stores data in physically contiguous blocks. The difference is in the use of indexes to locate records.
- To understand this method, to distinguish among three areas in disk storage: prime area, overflow area and index area.
- **The prime area** contains file records stored by key or ID numbers. All records are initially stored in the prime area.
- **The overflow area** contains records added to the file that cannot be placed in logical sequence in the prime area.
- **The index area** is more like a data dictionary. It contains keys of records and their locations on the disk. Almost similar to sequential method only that, an index is used to enable the computer to locate individual records on the storage media.
- Indexed-sequential organization reduces the magnitude of the sequential search and provides quick access for sequential and direct processing.
- For example, on a **magnetic drum**, records are stored sequential on the tracks. However, each record is assigned an index that can be used to access it directly.

ADVANTAGES of INDEXED FILES

- Quite easy to process.
- With proper selection of a key field, records in a large file can be searched and accessed in very quickly.
- Any field of the records can be used as the key. The key field can be numerical or alphanumerical.
- Indexed sequential organization reduces the magnitude of the sequential search and provides quick access for sequential and direct processing.
- Records can be inserted in the middle of the file.

DISADVANTAGES of INDEXED FILES

- Extra data structures have to be maintained (the COBOL run-time modules take care of these and it is not the programmers' concern). These extra data structures maintained on the disk can use up much disk space, especially for long key values.
- The indexed files have to be reorganized from time time to get rid of deleted records and improve performance that gets gradually decreased with addition of new records.
- It takes longer to search the index for data access or retrieval.
- Unique keys are required
- Periodic reorganization is required.

Database design:

- **Database** is a collection of related data and data is a collection of facts and figures that can be processed to produce information.
- Mostly data represents recordable facts. Data aids in producing information, which is based on facts. For example, if we have data about marks obtained by all students, we can then conclude about toppers and average marks.
- A **database management system** stores data in such a way that it becomes easier to retrieve, manipulate, and produce information.

Various Objectives of Database Management System

1) Mass Storage

DBMS can store a lot of data in it. So, for all the big firms, DBMS is really ideal technology to use. It can store thousands of records in it and one can fetch all that data whenever it is needed.

2)Controlled redundancy:

Redundant data occupies space and, therefore is wasteful. If versions of the same data are in different phases of updating, the system often gives conflicting information.

3) Removes Duplicity

If you have lots of data then data duplicity will occur for sure at any instance. DBMS guarantee it that there will be no data duplicity among all the records. While storing new records, DBMS makes sure that same data was not inserted before.

4)Ease of learning and use:

A major features of a user-friendly data base package is how easy it to learn and use. Related to this point is that a data base can be modified without interfering with establishing ways of using the data.

5)Data independence:

An important data base objective is changing hardware and storage procedures or adding new data without having to rewrite application programs.

6)Multiple Users Access

No one handles the whole database alone. There are lots of users who are able to access database. So, this situation may happen that two or more users are accessing database. They can change whatever they want, at that time DBMS makes it sure that they can work concurrently.

7) Data Protection

Information such as bank details, employee's salary details and sale purchase details should always be kept secured. Also, all the companies need their data secured from unauthorized use.

DBMS gives a master level security to their data. No one can alter or modify the information without the privilege of using that data.

8) Data Backup and recovery

Sometimes database failure occurs so there is no option like one can say that all the data has been lost. There should be a backup of database so that on database failure it can be recovered. DBMS has the ability to backup and recover all the data in database.

9) Everyone can work on DBMS

There is no need to be a master of programming language if you want to work on DBMS. Any accountant who is having less technical knowledge can work on DBMS. All the definitions and descriptions are given in it so that even a non-technical background person can work on it.

10) Accuracy and Integrity

The accuracy of a data base ensures that data quality and content remain constant. Integrity means your data is authentic and consistent. DBMS has various validity checks that make your data completely accurate and consistence.

11) Platform Independent

One can run DBMS at any platform. No particular platform is required to work on database management system.

12) More information at low cost:

Using, storing and modifying data at low cost are important.

LOGICAL AND PHYSICAL VIEWS OF DATA :

- Physical view refers to the way data are physically stored and processed in a database. On the other side, logical view is designed to suit the need of different users by representing data in a meaningful format. Another word, the logical view tells the users, in their term, what is in the database.
- A subschema is a subset of the schema and inherits the same property that a schema has. The plan (or scheme) for a view is often called subschema. Subschema refers to an application programmer's (user's) view of the data item types and record types, which he or she uses.
- Database changes over time when information is inserted or deleted. The collection of information stored in the database at a particular moment is called an instance of the database. The overall design of the database is called the database schema.
- A schema diagram, as shown above, displays only names of record types (entities) and names of data items (attributes) and does not show the relationships among the various files.

Instance of the schema

- The schema will remain the same while the values filled into it change from instant to instant. When the schema framework is filled in with data item values, it is referred as an instance of the schema. The data in the database at a particular moment of time is called a database state or snapshot, which is also called the current set of occurrences or instances in the database
- In other words, "the description of a database is called the database schema, which is specified during database design and is not expected to change frequently". A displayed schema is called a schema diagram.

Schema Diagram

A schema diagram displays only some aspects of a schema, such as the name. of record types and data items, and some types of constraints. Other aspects are not specified in the schema diagram. It does not specify the data type of each data item and the relationships among the various files.

Subschema

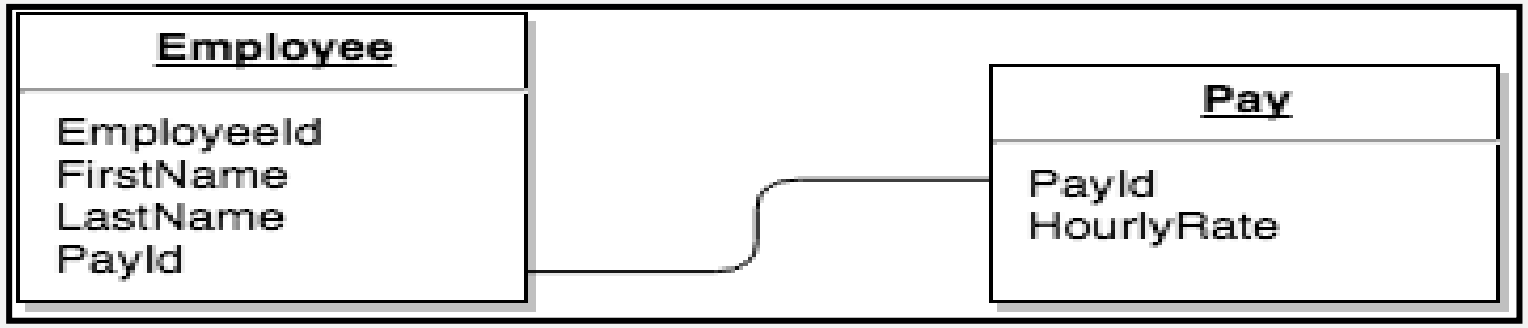
A subschema is a subset of the schema and inherits the same property that a schema has. The plan (or scheme) for a view is often called subschema. Subschema refers to an application programmer's (user's) view of the data item types and record types, which he or she uses. It gives the users a window through which he or she can view only that part of the database, which is of interest to him. Therefore, different application programs can have different view of data.

There are 3 types of relationships in relational database design. They are:

- One-to-One
- One-to-Many (or Many-to-One)
- Many-to-Many

One-to-One

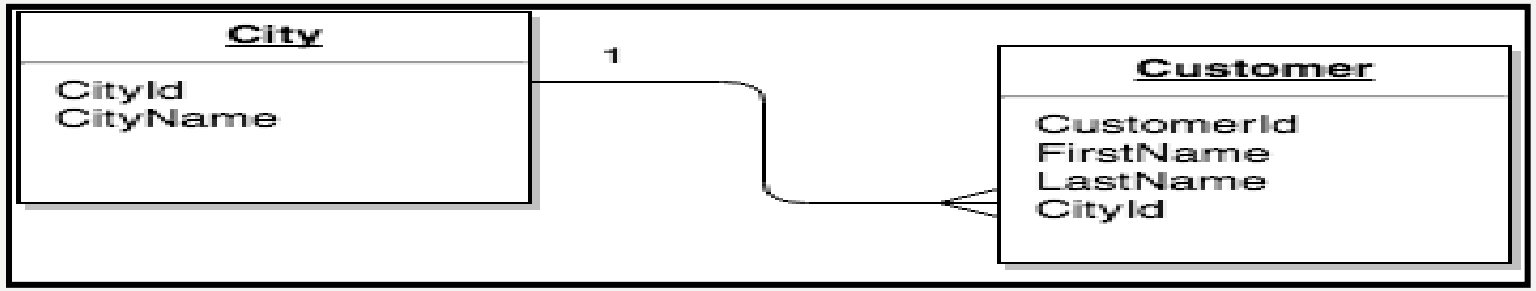
A row in table A can have only one matching row in table B, and vice versa.



- This is not a common relationship type, as the data stored in table B could just have easily been stored in table A. However, there are some valid reasons for using this relationship type. A one-to-one relationship can be used for security purposes, to divide a large table, and various other specific purposes.
- In the above example, we could just as easily have put an Hourly Rate field straight into the Employee table and not bothered with the Pay table. However, hourly rate could be sensitive data that only certain database users should see. So, by putting the hourly rate into a separate table, we can provide extra security around the Pay table so that only certain users can access the data in that table.

One-to-Many (or Many-to-One)

This is the most common relationship type. In this type of relationship, a row in table A can have many matching rows in table B, but a row in table B can have only one matching row in table A.



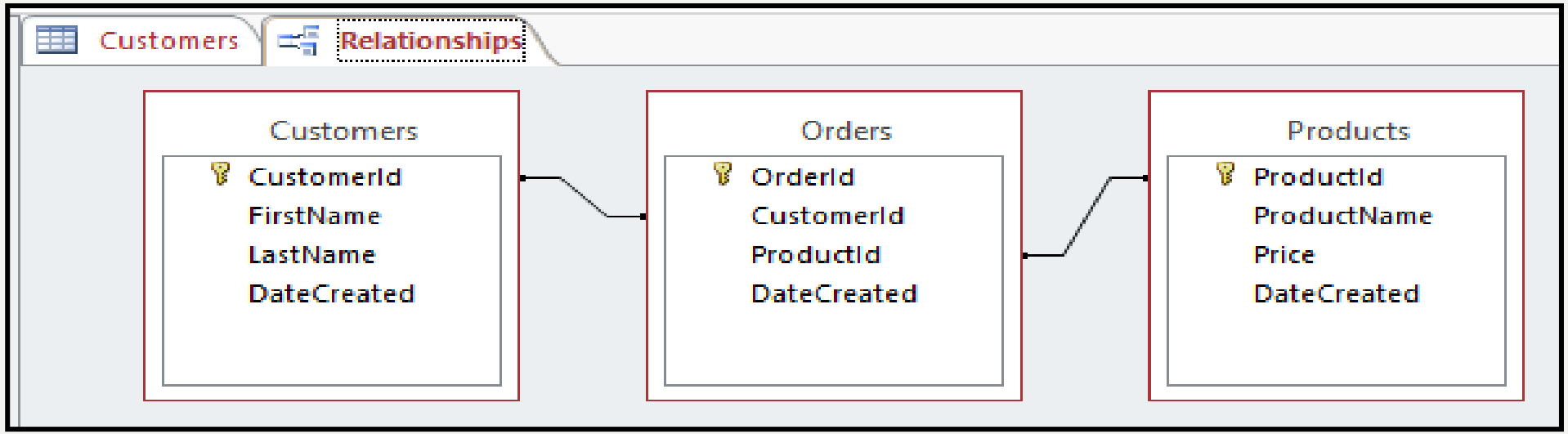
Example of one-to-many relationship.

One-to-Many relationships can also be viewed as Many-to-One relationships, depending on which way you look at it.

In the above example, the Customer table is the “many” and the City table is the “one”. Each customer can only be assigned one city,. One city can be assigned to many customers.

Many-to-Many

- In a many-to-many relationship, a row in table A can have many matching rows in table B, and vice versa.
- A many-to-many relationship could be thought of as two one-to-many relationships, linked by an intermediary table.
- The intermediary table is typically referred to as a “junction table” (also as a “cross-reference table”). This table is used to link the other two tables together. It does this by having two fields that reference the primary key of each of the other two tables.
- The following is an example of a many-to-many relationship:



- This is the Relationships tab that is displayed when you create a relationship Microsoft Access. In this case, a many-to-many relationship has just been created. The Orders table is a junction table that cross-references the Customers table with the Products table.
- So in order to create a many-to-many relationship between the Customers table and the Products table, we created a new table called Orders.

Normalization

If a database design is not perfect, it may contain anomalies, which are like a bad dream for any database administrator. Managing a database with anomalies is next to impossible.

- Normalization is the process of organizing the data in the database.
- Normalization is used to minimize the redundancy from a relation or set of relations. It is also used to eliminate the undesirable characteristics like Insertion, Update and Deletion Anomalies.
- Normalization divides the larger table into the smaller table and links them using relationship.
- The normal form is used to reduce redundancy from the database table.

Update anomalies − If data items are scattered and are not linked to each other properly, then it could lead to strange situations. For example, when we try to update one data item having its copies scattered over several places, a few instances get updated properly while a few others are left with old values. Such instances leave the database in an inconsistent state.

Deletion anomalies − We tried to delete a record, but parts of it was left undeleted because of unawareness, the data is also saved somewhere else.

Insert anomalies − We tried to insert data in a record that does not exist at all.

Normalization is a method to remove all these anomalies and bring the database to a consistent state.

First Normal Form

First Normal Form is defined in the definition of relations (tables) itself. This rule defines that all the attributes in a relation must have atomic domains. The values in an atomic domain are indivisible units.

- A relation will be 1NF if it contains an atomic value.
- It states that an attribute of a table cannot hold multiple values. It must hold only single-valued attribute.
- First normal form disallows the multi-valued attribute, composite attribute, and their combinations.

Course	Content
Programming	Java, c++
Web	HTML, PHP, ASP

We re-arrange the relation (table) as below, to convert it to First Normal Form.

Course	Content
Programming	Java
Programming	c++
Web	HTML
Web	PHP
Web	ASP

Each attribute must contain only a single value from its pre-defined domain.

Second Normal Form

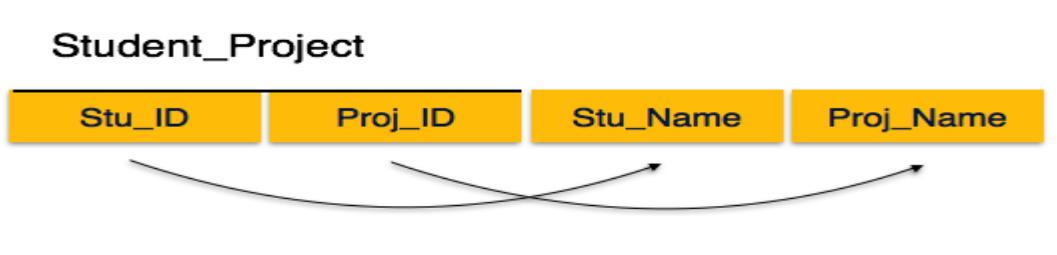
Before we learn about the second normal form, we need to understand the following −

- In the 2NF, relational must be in 1NF.
- In the second normal form, all non-key attributes are fully functional dependent on the primary key

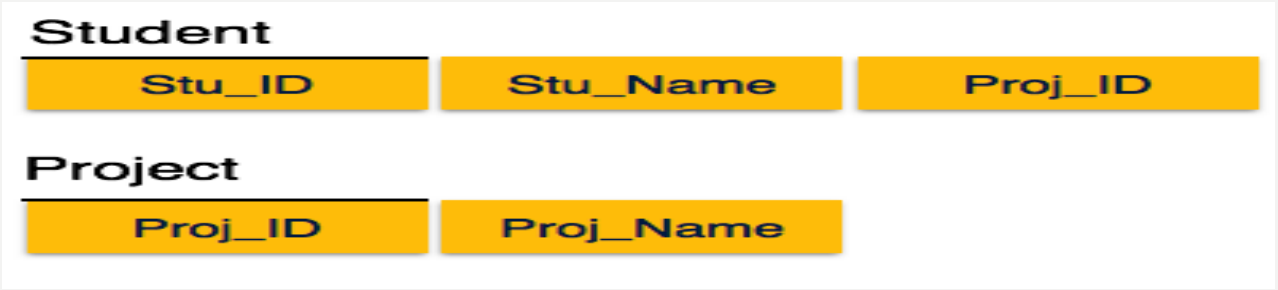
Prime attribute − An attribute, which is a part of the candidate-key, is known as a prime attribute.

Non-prime attribute − An attribute, which is not a part of the prime-key, is said to be a non-prime attribute.

If we follow second normal form, then every non-prime attribute should be fully functionally dependent on prime key attribute. That is, if X → A holds, then there should not be any proper subset Y of X, for which Y → A also holds true..



We see here in Student_ Project relation that the prime key attributes are Stu_ID and Proj_ID. According to the rule, non-key attributes, i.e., Stu_Name and Proj_Name must be dependent upon both and not on any of the prime key attribute individually. But we find that Stu_Name can be identified by Stu_ID and Proj_Name can be identified by Proj_ID independently. This is called partial dependency, which is not allowed in Second Normal Form.



Third Normal Form

For a relation to be in Third Normal Form, it must be in Second Normal form and the following must satisfy −No non-prime attribute is transitively dependent on prime key attribute.

•A relation will be in 3NF if it is in 2NF and not contain any transitive partial dependency.
•3NF is used to reduce the data duplication. It is also used to achieve the data integrity.
•If there is no transitive dependency for non-prime attributes, then the relation must be in third normal form.

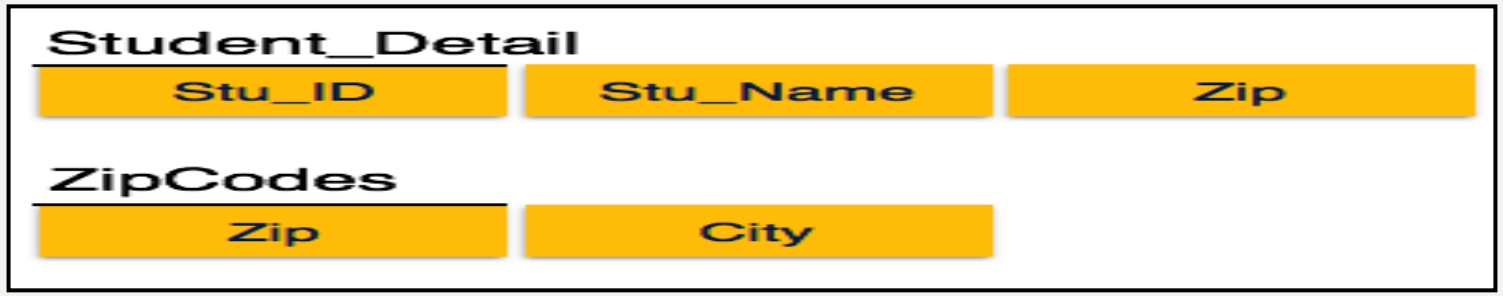
A relation is in third normal form if it holds atleast one of the following conditions for every non-trivial function dependency X → Y.

1.X is a super key.
2.Y is a prime attribute, i.e., each element of Y is part of some candidate key.

- For any non-trivial functional dependency, X → A, then either −
  - X is a superkey or,
  - A is prime attribute.

❖ We find that in the above Student detail relation, Stu_ID is the key and only prime key attribute. We find that City can be identified by Stu_ID as well as Zip itself. Neither Zip is a superkey nor is City a prime attribute. Additionally, Stu_ID → Zip → City, so there exists transitive dependency.

❖ To bring this relation into third normal form, we break the relation into two relations as follows −



Role and responsibilities of Database Administrator:

Software installation and Maintenance

❖ A DBA often collaborates on the initial installation and configuration of a new Oracle, SQL Server etc. database.

❖ The system administrator sets up hardware and deploys the operating system for the database server, then the DBA installs the database software and configures it for use. As updates and patches are required, the DBA handles this on-going maintenance.

❖ And if a new server is needed, the DBA handles the transfer of data from the existing system to the new platform.

2. Data Extraction, Transformation, and Loading

Known as ETL, data extraction, transformation, and loading refers to efficiently importing large volumes of data that have been extracted from multiple systems into a data warehouse environment.

This external data is cleaned up and transformed to fit the desired format so that it can be imported into a central repository.

3.Specialized Data Handling

Today’s databases can be massive and may contain unstructured data types such as images, documents, or sound and video files.

Managing a very large database (VLDB) may require higher-level skills and additional monitoring and tuning to maintain efficiency.

4. Database Backup and Recovery

DBAs create backup and recovery plans and procedures based on industry best practices, then make sure that the necessary steps are followed. Backups cost time and money, so the DBA may have to persuade management to take necessary precautions to preserve data.

- System admin or other personnel may actually create the backups, but it is the DBA's responsibility to make sure that everything is done on schedule.
- In the case of a server failure or other form of data loss, the DBA will use existing backups to restore lost information to the system.
- Different types of failures may require different recovery strategies, and the DBA must be prepared for any eventuality. With technology change, it is becoming ever more typical for a DBA to backup databases to the cloud, Oracle Cloud for Oracle Databases and MS Azure for SQL Server.

5.Security:

- A DBA needs to know potential weaknesses of the database software and the company's overall system and work to minimize risks. No system is one hundred per cent immune to attacks, but implementing best practices can minimize risks.
- In the case of a security breach or irregularity, the DBA can consult audit logs to see who has done what to the data. Audit trails are also important when working with regulated data.

6. Authentication

- Setting up employee access is an important aspect of database security. DBAs control who has access and what type of access they are allowed.
- For instance, a user may have permission to see only certain pieces of information, or they may be denied the ability to make changes to the system.

7. Capacity Planning

- The DBA needs to know how large the database currently is and how fast it is growing in order to make predictions about future needs. Storage refers to how much room the database takes up in server and backup space. Capacity refers to usage level.
- If the company is growing quickly and adding many new users, the DBA will have to create the capacity to handle the extra workload.

8. Performance Monitoring

Monitoring databases for performance issues is part of the on-going system maintenance a DBA performs. If some part of the system is slowing down processing, the DBA may need to make configuration changes to the software or add additional hardware capacity.

Many types of monitoring tools are available, and part of the DBA's job is to understand what they need to track to improve the system. 3rd party organizations can be ideal for outsourcing this aspect, but make sure they offer modern DBA support.

9. Database Tuning

Performance monitoring shows where the database should be tweaked to operate as efficiently as possible. The physical configuration, the way the database is indexed, and how queries are handled can all have a dramatic effect on database performance.

With effective monitoring, it is possible to proactively tune a system based on application and usage instead of waiting until a problem develops.

10. Troubleshooting

DBAs are on call for troubleshooting in case of any problems.

Whether they need to quickly restore lost data or correct an issue to minimize damage, a DBA needs to quickly understand and respond to problems when they occur.


Thank You