

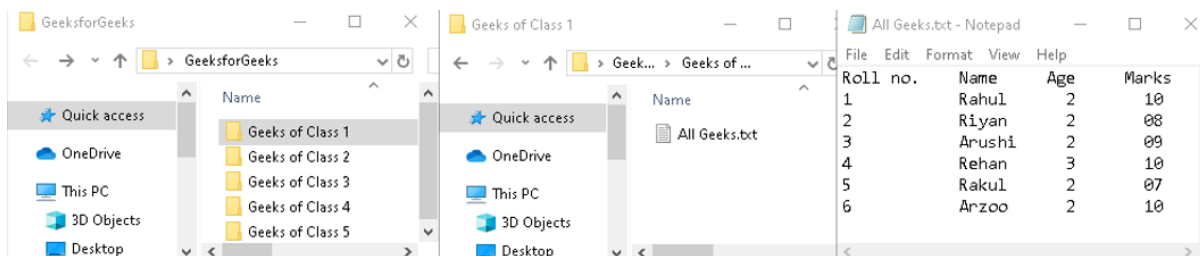
Overview of File Processing

In earlier days, data was stored manually, using pen and paper but after the computer was discovered, the same task could be done by using files.

A computer **File** is a resource that uniquely records data, in a storage device on a computer.

There are various formats in which data can be stored. e.g. Text files can be stored in **.txt format** while pictures can be stored in **.png format** etc.

In the case of computer files, data about data (**metadata**) can be stored in different lines, separated by spaces, commas, or tab to resemble tables. Each file is placed in relevant folders for ease of access.



In Computer Science, File Processing System (FPS) is a way of storing, retrieving, and manipulating data that is present in various files. Files are used to store various documents.

Data Design Concepts

Before constructing an information system, a systems analyst must understand basic design concepts, including data structures and the characteristics of the file processing and database systems, including Web-based database design.

The data storage function manages how data is stored and handled by programs that run the system

Goals of data storage design

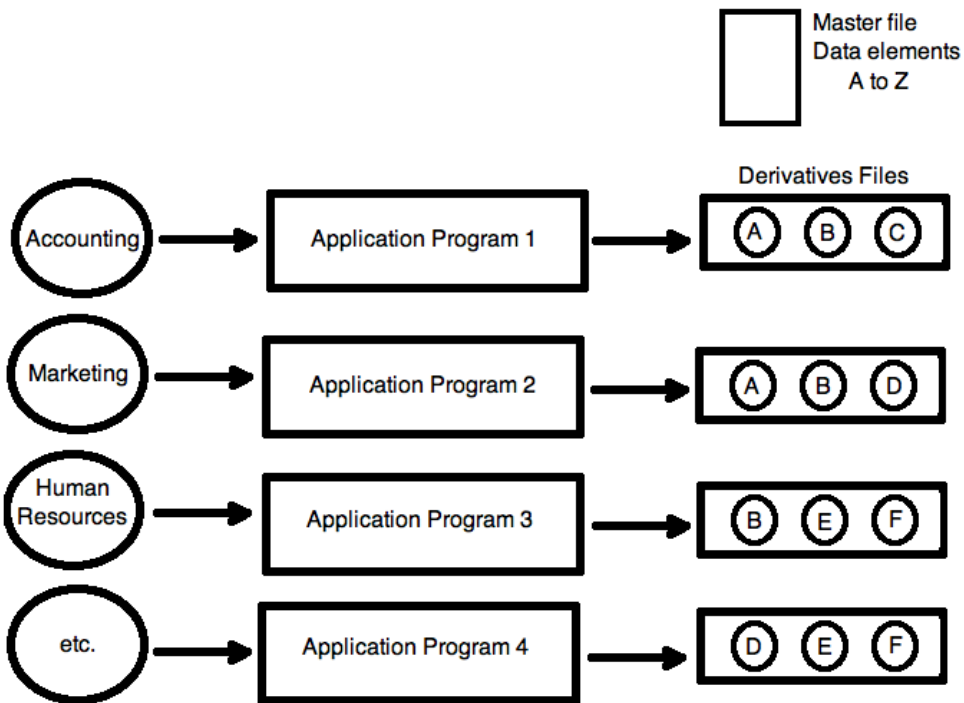
- Efficient data retrieval (good response time)
- Access to the information users need

Traditional File Processing System

File processing systems were an early attempt to computerize the manual filing system that we are all familiar with. A file system is a method for storing and organizing computer files and the data they contain to make it easy to find and access them.

In OS, a **file system** is also a process that manages how and where data on a storage disk, typically a hard disk drive (HDD), is stored, accessed and managed.

In the traditional approach, we used to store information in flat files that are maintained by the file system under the operating system's control. Here, flat files are files containing records having no structured relationship among them.



Characteristics of File Processing System

Here is the list of some important characteristics of a file processing system:

It is a group of files storing data of an organization.

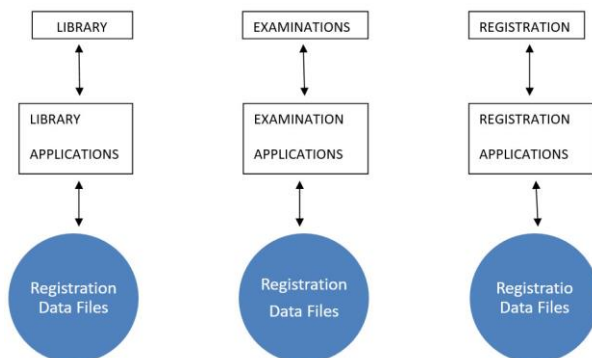
- Each file is independent of one another.
- Each file is called a flat-file.
- Each file contained and processed information for one specific function, such as accounting or inventory.
- The physical implementation and access procedures are written into the database application; therefore, physical changes resulted in intensive rework on the part of the programmer.
- As systems became more complex, file processing systems offered little flexibility, presented many limitations, and were difficult to maintain.

Advantages of File Processing System:

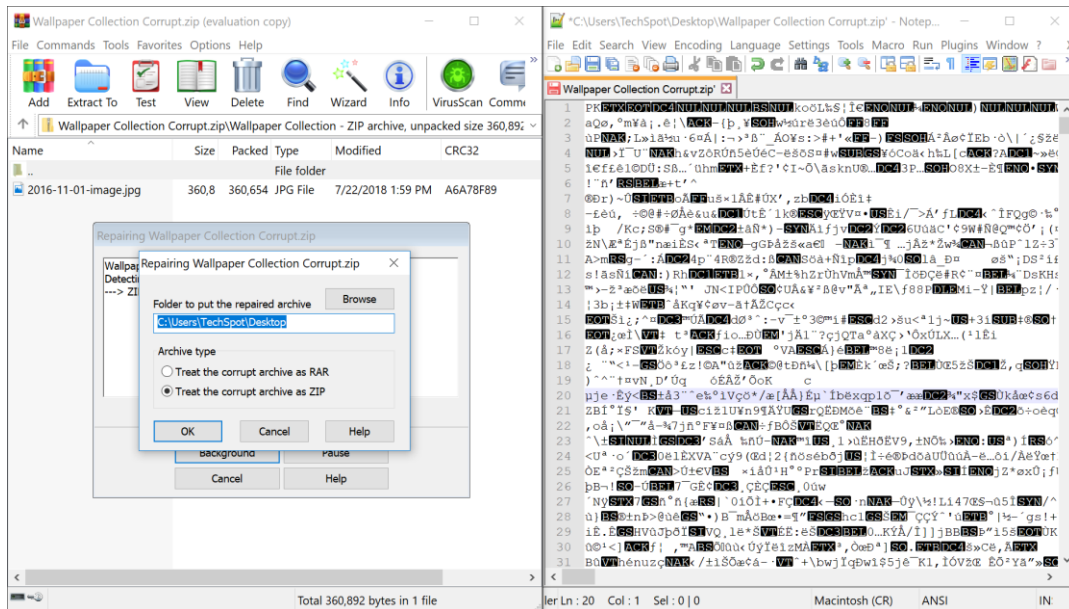
- **Cost friendly** – There is a very minimal to no setup and usage fee for File Processing System. (In most cases, free tools are inbuilt in computers.)
- **Easy to use** – File systems require very basic learning and understanding, hence, can be easily used.
- **High scalability** – One can very easily switch from smaller to larger files as per his needs.

Disadvantages of File Processing System:

- **Slow access time** – Direct access of files is very difficult and one needs to know the entire hierarchy of folders to get to a specific file. This involves a lot of time.



- **Presence of redundant data** – The same data can be present in two or more files which takes up more disc space.
- **Difficulty in the recovery of corrupt data** – Recovery or backup of lost and corrupt data is nearly impossible in the case of the File Processing System.



Database Terminologies

Databases and database systems are an essential component of life in modern society: most of us encounter several activities every day that involve some interaction with a database

- Databases and database technology have a major impact on the growing use of computers. It is fair to say that databases play a critical role in almost all areas where computers are used, including business, electronic commerce, engineering, medicine, genetics, law, education, and library science
- The word database is so commonly used that we must begin by defining what a database is. Our initial definition is quite general.
- A database is a collection of related data. By data, we mean known facts that can be recorded and that have implicit meaning. For example, consider the names, telephone numbers, and addresses of the people you know.

A database has the following implicit properties:

- A database represents some aspect of the real world, sometimes called the miniworld or the universe of discourse (UoD).
- A database is a logically coherent collection of data with some inherent meaning. A random assortment of data cannot correctly be referred to as a database.
- A database is designed, built, and populated with data for a specific purpose.

It has an intended group of users and some preconceived applications in which these users are interested.

1. **Database Administrators** - responsible for authorizing access to the database, coordinating and monitoring its use, and acquiring software and hardware resources as needed.
2. **Database Designers** - responsible for identifying the data to be stored in the database and for choosing appropriate structures to represent and store this data.

Examples of large databases:

Amazon.com, Canadian Census, The Bay's product inventory, data collection underlying Quest

Database management system (DBMS)

- Collection of programs
- Enables users to create and maintain a database
- Allows multiple users and programs to access and manipulate the database concurrently

- Provides protection against unauthorized access and manipulation
- Provides means to evolve database and program behavior as requirements change over time

Defining a database - Specifying the data types, structures, and constraints of the data to be stored. Uses a Data Definition Language

Meta-data - Database definition or descriptive information • Stored by the DBMS in the form of a database catalog or data dictionary

Phases for designing a database:

1. Requirements specification and analysis
2. Conceptual design, e.g., using the Entity-Relationship model
3. Logical design, e.g., using the relational model
4. Physical design

Query - Interaction causing some data to be retrieved • uses a Query Language ♣ Manipulating a database

Application program - Accesses database by sending queries and updates to DBMS

Transaction - An atomic unit of queries and updates that must be executed as a whole

DBMS SCHEMATIC

An example of a **UNIVERSITY** database for maintaining information concerning students, courses, and grades in a university environment.

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

COURSE

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

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

Characteristics of the database

The main characteristics of the database approach versus the file-processing approach are the following:

- Self-describing nature of a database system
- Separation of Data and Applications
- Support for multiple views of the data
- Data Integrity
- Data Persistence
- Sharing of data and multiuser transaction processing