

Democratic and Popular Republic of Algeria

Ministry of Higher Education and Scientific Research



*Ecole supérieure en sciences et technologies de
l'informatique et du numérique*

General Information on Files

Presented by : Dr. Daoudi Meroua

Academic year: 2024/2025

Plan

- ❑ Basic Concepts
- ❑ File Usage Characteristics
- ❑ File Typology
- ❑ Fundamental Operations on Files
- ❑ Memories
- ❑ Physical Files and Logical Files
- ❑ Physical Record and Logical Record
- ❑ Blocking Factor and Its Importance
- ❑ Conclusion

Introduction

A file is a collection of data stored in secondary memory; this data can be processed by programs. Programs are also stored in the form of files. .

Introduction

Program Files: These are files that contain the instructions of a program. A program file consists of:

Source File: This file has the extension corresponding to the language used in the program, such as .c, .java, .pas, etc.

Executable File: These are files that can be executed, typically with extensions like .exe, .drv, or .vbx.

Introduction

Data Files: These are files that group the data that a program may use and/or the results it produces. Data files are dynamic, meaning that data can be modified, deleted (erased), added (inserted), or simply viewed.

Example:

The file `student.txt`, which contains a set of characters created by a program or a text editor. The contents of the file can be modified (insertion, deletion, modification) or consulted.

Introduction

The files of interest here are structured data files (such as those for people, works, products, etc.) created by the user for the management of a specific application.

Basic Concepts

The information processed by the computer is represented in binary, the only form that can be understood by the machine. Thus, any information is converted into a sequence of bits (0 or 1). This information is organized into:

- Characters
- Fields
- Records
- Files

Basic Concepts

Character

A character is a grouping of bits used to represent an alphanumeric character (0, 1, 2, ... a, b, c, ...) or a special character (&, !, \$, ...) in any encoding.

A character is the smallest technologically significant grouping of bits.

Basic Concepts

Field

A field refers to a grouping of successive characters that represent information accessible by a process within a program. To distinguish between different fields, identifiers are assigned to them.

Example:

```
var nombre1, nombre2, somme: integer;
```

A field can be further decomposed into other fields.

Example:

The field "date" can be decomposed into three fields: day, month, year.

Basic Concepts

Record

When there are fields that contain information related to a single subject, we refer to it as a record.

Example:

Number: 001 Name: Ait Ahmed First Name: Fateh Address: Oran

A record is a collection of information contained in fields that relate to the same subject.

Basic Concepts

Note:

It is observed that the fields of two different records can be identical, which may lead to ambiguity. To avoid this situation, there must be information that has a unique value for each student: in our example, the student number allows us to distinguish between students. This information constitutes what we call a key or identifier.

Basic Concepts

Regardless of the physical medium used to store it, a file is a set of information of the same nature that describes individuals or objects (concrete or abstract) possessing common characteristics.

Example:

The Students file groups the information (Name, First Name, etc.) concerning the students of a given educational institution.

File Usage Characteristics

A file is created to be used for the management of any application, to perform calculations, etc. Depending on the importance of the data it contains, it undergoes varying levels of manipulations that determine the file's usage characteristics, namely:

- ➔ The activity of a file
- ➔ The volume of a file
- ➔ The growth of a file

File Usage Characteristics

The activity of a file characterizes all the manipulations performed on the file. It is defined by the following characteristics:

➔ **Rate of access**

➔ **Rate of update**

➔ **Stability of the file**

File Usage Characteristics

✓ **Rate of Access:** This refers to the ratio between the number of records accessed (or modified) and the total number of records in the file over a certain period.

$RA = \text{number of records accessed (or modified)} / \text{total number of records}$

Rate of Update: This is relative to a given period. It expresses the relative number of new records that are inserted into the file.

File Usage Characteristics

- ✓ **File stability:** It refers to a given period. A file is considered stable during a period if the number of records created is approximately equal to the number of records deleted

File Usage Characteristics

The volume or size of a file:

It refers to the number of characters contained in the file. This is a very important characteristic for the future use of the file (physical implementation, estimation of file manipulation time).

File growth:

It refers to the number of records created in relation to the number of records deleted. It is considered negative when the number of records deleted exceeds the number of records created.

File Usage Characteristics

Several types of files can be distinguished according to:

- ➔ The nature of the information they contain.
- ➔ Their lifespan.
- ➔ The type of storage medium used.
- ➔ The organization of the information

File Usage Characteristics

Types of files based on the nature of the information:

A file can contain two types of information: data or programs.

Depending on the case, we refer to a data file or a program file.

Example:

The files students and modules are data files.

The file calculation is a program file used to calculate students' average grades.

File Usage Characteristics

Types of Files Based on Their Lifespan

Permanent Files:

Definition: These files contain crucial, long-term data used by an application.

Example: In student management, the "student record" file is a permanent file, as its contents are not frequently updated but remain critical for long-term reference.

File Usage Characteristics

Types of Files Based on Their Lifespan

Movement Files :

Definition: Temporary files used to update permanent files. They have a short lifespan and are deleted after their purpose is fulfilled.

Example: A file containing newly registered students each semester. Once their details are added to the permanent student file, this movement file is no longer needed.

File Usage Characteristics

Types of Files Based on Their Lifespan

Maneuver Files

Definition: These files are used when there is insufficient memory to handle large datasets. They exist only as long as the associated process runs.

Example: If a school system processes many students' exam scores, a maneuver file may temporarily hold the data of those students who are due for makeup exams.

File Usage Characteristics

Types of Files Based on Their Lifespan

Intermediate Files:

Definition: These files store the results of a process temporarily. They can be used in later stages of the same process or by other processes.

Example: A file storing semester-end results, which might later be used to assign students to internships based on their performance.

File Usage Characteristics

Types of Files Based on Their Lifespan

Archive and Historical Files:

Definition: These files keep track of past operations or data. The difference between archive and historical files usually lies in the frequency of their updates.

Example: At the end of every semester, the results are stored in a historical file, which is periodically updated. An archive file might store older data that is no longer modified but kept for record-keeping purposes.

File Usage Characteristics

Types of files based on the storage medium:

Some file characteristics are closely related to the type of storage medium, particularly the method of data access.

Example:

For files stored on magnetic tape, only sequential access is possible.

On a magnetic disk, data access can be either sequential or direct.

File Usage Characteristics

Types of files based on file organization:

The organization adopted for a file is one of its most important characteristics, as it defines how to access the information it contains.

Example:

There are mainly three types of file organization:

- ➔ Sequential organization
- ➔ Indexed organization
- ➔ Random organization
- ➔ Tree based organisation