

STRUCTURE DE FICHIERS & STRUCTURE DE DONNÉES

CHAPITRE I GÉNÉRALITÉS SUR LES FICHIERS

Contenu de chapitre

1. Mémoire centrale vs. Mémoire secondaire
2. Définitions et concepts de base :
 1. Fichier, enregistrement, zone
 2. Fichier physique vs. Fichier logique
 3. Fichier statique vs. Fichier dynamique
3. Caractéristiques d'un fichier
4. Topologie des fichiers
5. Opérations de bases sur les fichiers



Mémoire centrale vs. Mémoire secondaire

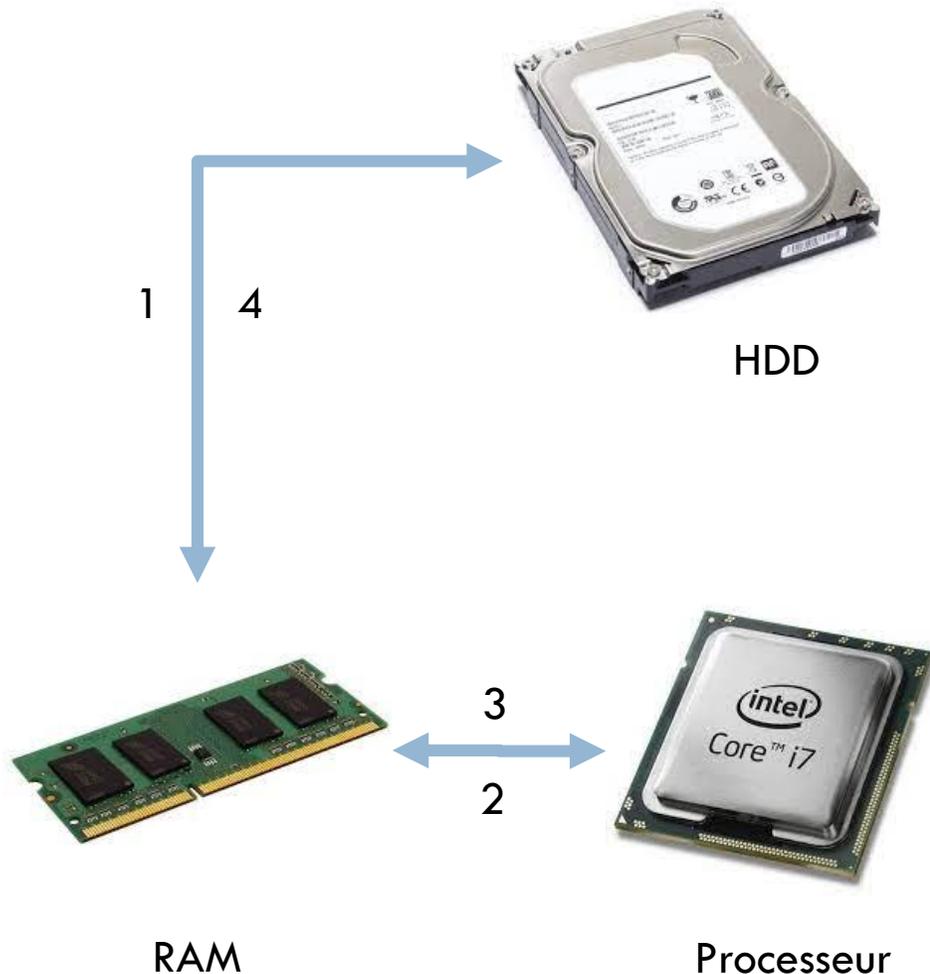
Mémoire secondaire

- La mémoire secondaire est la mémoire de stockage de l'ordinateur. Elle conserve les données même si l'ordinateur est éteint (**non-volatile**).
- Elle est moins sollicitée par le CPU et contient les fichiers et données des programmes qui ne sont pas utilisés par des programmes en cours d'exécution.
- Il existe plusieurs types de mémoire secondaire classées selon la possibilité de les modifier :
 - ▣ ROM (*Read-Only Memory*) : n'est pas effaçable et utile pour stocker les instructions de démarrage d'un ordinateur.
 - ▣ EEPROM (*Electrically Erasable Programmable Read Only Memory*) : la suppression et la modification des données est possible – Dique dur, Clé USB, Disque optique (CD ROM, DVD), Carte mémoire (SD, microSD)

Mémoire centrale

- Elle est aussi appelée mémoire vive, primaire ou RAM pour *Random Access Memory*
- La mémoire vive est la mémoire de travail de l'ordinateur. Pour exécuter un programme, l'ordinateur doit copier les données depuis une mémoire secondaire vers la mémoire vive.
- Elle est caractérisée par sa **volatilité** : les données s'effacent lors de l'extinction du système.
- Elle contient les données des processus qui s'exécutent sur l'ordinateur et est constamment sollicitée par le processeur pour traiter ces données.

Mémoire centrale vs. Mémoire secondaire



Mémoire centrale	Mémoire secondaire
Accès rapide	Accès lent
Volatile	Non-volatile
Coûteuse	Bon marché
Espace en Go	Espace en To

Définitions et concepts de base

Fichier, Enregistrement, Zone

Bit = Binary digit = nombre binaire

- Un bit représente une cellule électronique soumise à un champ électrique ou électromagnétique, qui peut saisir l'un des états **0** (pas de courant électrique) ou **1** (il y a du courant électrique).
- Avec un seul bit on ne peut pas modéliser beaucoup d'informations, mais avec plusieurs bits on peut coder plus d'informations.

Le bit est l'unité de base avec laquelle on peut coder n'importe quelle information du monde physique.

Cas de 1 bit

0
1

Cas de 2 bits

0	0
0	1
1	0
1	1

Cas de 3 bits

0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Cas de 4 bits

0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Bit: Codage des textes

- Un texte écrit est constitué de:
 - ▣ Alphabets = 26 X 2
 - ▣ Des chiffres qui sont au nombre de 10
 - ▣ Symboles mathématiques: (+, -, *, /, <, >)
 - ▣ caractères de ponctuations (“.”, “;”, “?”, “!”)
 - ▣ Caractères spéciaux (\$, £, &, §, @).
 - ▣ En résumé, une étude exhaustive amène à 256 caractères possibles → 8 bits
- Une table de correspondance (caractère/code binaire) est associée pour le codage et le décodage des textes.

Caractère	Code Binaire	Caractère	Code Binaire	Caractère	Code Binaire
espace	00100000	H	01001000	c	01100011
!	00100001	I	01001001	d	01100100
"	00100010	J	01001010	e	01100101
#	00100011	K	01001011	f	01100110
\$	00100100	L	01001100	g	01100111
%	00100101	M	01001101	h	01101000
&	00100110	N	01001110	i	01101001
'	00100111	O	01001111	j	01101010
(00101000	P	01010000	k	01101011
)	00101001	Q	01010001	l	01101100
0	00110000	R	01010010	m	01101101
1	00110001	S	01010011	n	01101110
2	00110010	T	01010100	o	01101111
3	00110011	U	01010101	p	01110000
4	00110100	V	01010110	q	01110001
5	00110101	W	01010111	r	01110010
6	00110110	X	01011000	s	01110011
7	00110111	Y	01011001	t	01110100
8	00111000	Z	01011010	u	01110101
9	00111001	[01011011	v	01110110
A	01000001	/	01011100	w	01110111
B	01000010]	01011101	x	01111000
C	01000011	^	01011110	y	01111001
D	01000100	-	01011111	z	01111010
E	01000101	`	01100000	{	01111011
F	01000110	a	01100001		01111100
G	01000111	b	01100010	}	01111101

Bonjour = 01000010 01101111 01101110 01101001 01101111 01110101 01110010

Bit: Codage des nombres

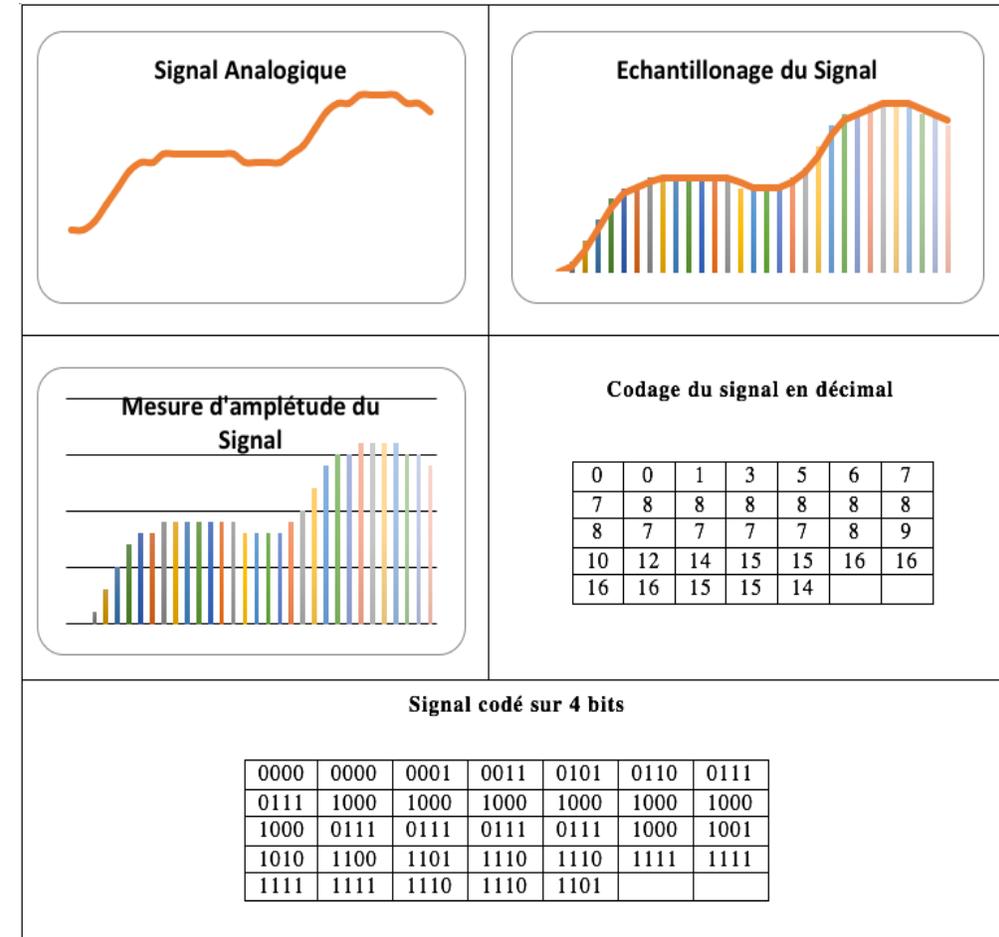
- Un nombre est une concaténation arbitraire d'une partie ou de la totalité des 10 chiffres disponibles (0..9)
 - ▣ Le nombre 123 est composé syntaxiquement de 1 puis de 2 puis de 3.
 - ▣ Sémantiquement ce nombre est obtenu par la multiplication des puissance de 10 comme suit: $123 = 1*100 + 2*10 + 3 = 1*10^2 + 2*10^1 + 3*10^0$
 - ▣ La multiplication des puissance de 10 vient du principe que nous utilisons un codage en base 10 (à base de 10 chiffres).
 - ▣ Le même principe a été adopté pour les machines où un codage en base 2 (deux chiffres uniquement sont autorisés 0 et 1). Un nombre sur une machine peut être syntaxiquement représenté par une séquence de 0 et de 1, mais sémantiquement il est représenté par des multiples de puissance de 2.

$$123 =_{\text{binaire}} 1111011$$

$$= 1*2^6 + 1*2^5 + 1*2^4 + 1*2^3 + 0*2^2 + 1*2^1 + 1*2^0$$

Bit: Codage des sons

- Un son est une pression de l'air dans un intervalle de temps. Il est modélisé par un signal analogique continu en fonction du temps.
- Pour coder un signal, il faut :
 - ▣ **Échantillonnage:** découper le signal en de petits morceaux (échantillons) égaux, chacun correspond à une partie du signal dans un intervalle de temps fixe.
 - ▣ **Mesure de l'amplitude:** associer à chaque échantillon le niveau du signal. Ces valeurs sont des nombres, donc un signal peut être systématiquement représenté par une suite de nombres. Ces nombres peuvent être codés en binaire selon le processus de codage des nombres.



Caractère

- Un caractère est le regroupement de 8 bits, appelé aussi Octet (byte en anglais).
- Le caractère ou l'octet (byte) est l'unité de mesure de l'information.
- Différentes autres unités ont été introduites. Ces unités sont principalement utilisées pour exprimer la taille de l'espace nécessaire pour le stockage de l'information.
 - ▣ Clé USB qui stocke entre 64 Mo et 32 Go
 - ▣ CD contient environ 600 Mo
 - ▣ DVD qui peut stocker jusqu'à 17 Go
 - ▣ Disques durs se comptent en To.

Unité de mesure	Taille	En octets
1 Octet (O)	8 bits ($8 = 2^3$)	1 octet
1 Kilo-octet (Ko)	1024 octets	2^{10} octets
1 Mega-octet (Mo)	1024 Ko	2^{20} octets
1 Giga-octet (Go)	1024 Mo	2^{30} octets
1 Téra-octet (To)	1024 Go	2^{40} octets
1 Péta-octet (Po)	1024 To	2^{50} octets
1 Exa-octet (Eo)	1024 Po	2^{60} octets
1 Zetta-octet (Zo)	1024 Eo	2^{70} octets
1 Yotta-octet (Yo)	1024 Zo	2^{80} octets

Zone

- La zone est le regroupement successif de plusieurs caractères (8 bits).
- La zone représente une information accessible par un traitement.
- Pour distinguer entre les différentes zones on les associe des identificateurs.
- La taille ou la longueur d'une zone est le nombre de caractère de cette zone.
- La nature des caractères de la zone définit le type de cette dernière.

- Exemples:

- Numéro de sécurité social,
- Nom,
- Prénom,
- Date de naissance,
- Adresse,

Types en langage C	Taille
int, short int	2 Octets
long, float	4 Octets
double	8 Octets
long double	10 Octets
char	1 Octet

Enregistrement

- Un enregistrement est le regroupement de plusieurs zones qui rapportent à un seul sujet.
- Un enregistrement est un ensemble de données (informations) stockées dans un fichier.
- Pour identifier les enregistrements d'une façon unique une zone appelée clé est associée où sa valeur est unique pour chaque enregistrement.
- Exemples:
 - ▣ Les informations décrivant les étudiants dans une université : Matricule, Nom , Prénom, Date de naissance, Moyenne.
 - ▣ Les informations décrivant les véhicules d'un parc automobile : matricule, marque, type, couleur.

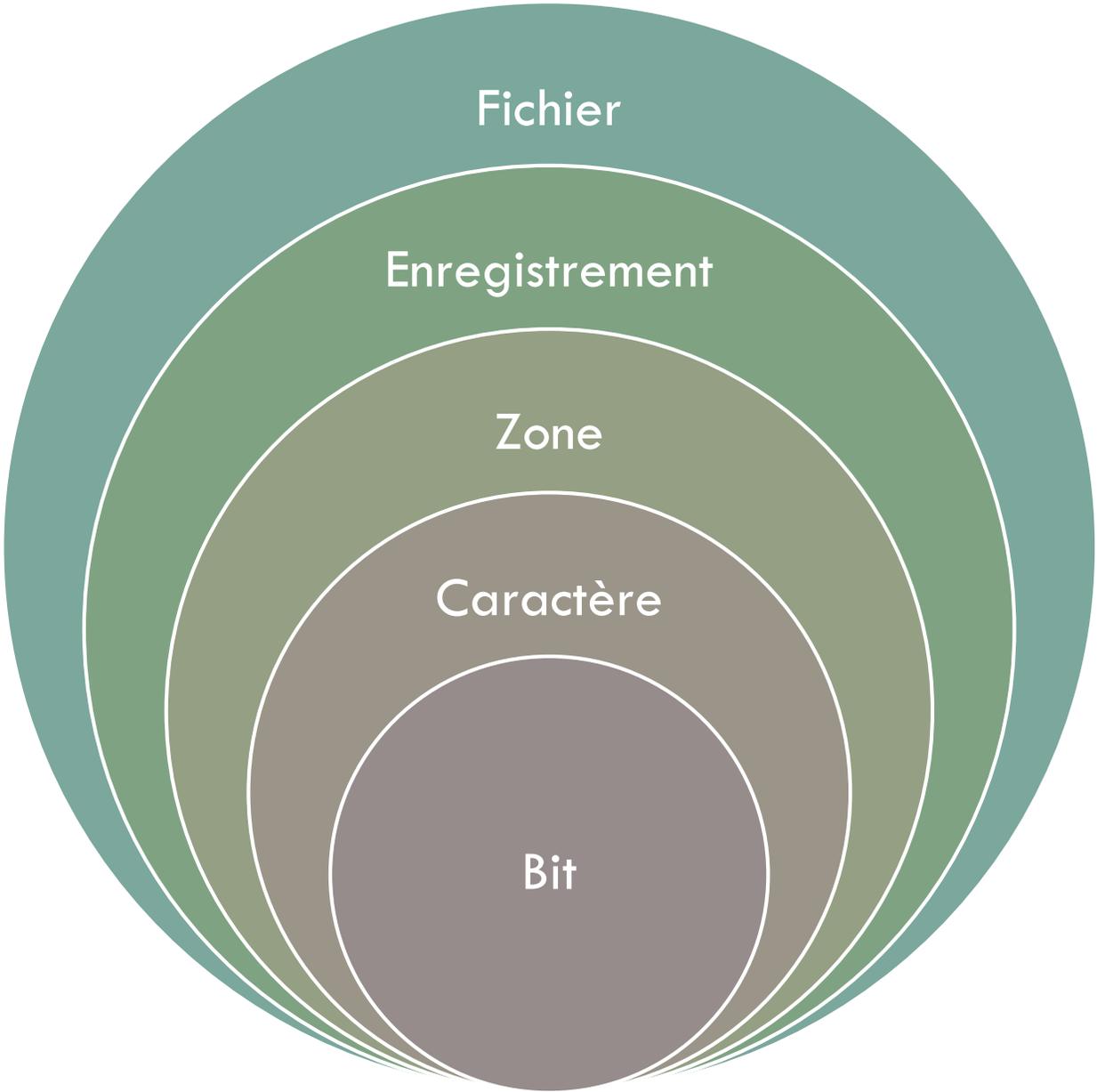
Les enregistrement en langage C

- Un enregistrement (**structure** en langage C), est une variable complexe qui permet de désigner sous un seul nom un ensemble de valeurs pouvant être de type différent.
- Initialisation d'un enregistrement:

```
1 ▸ struct nom_enregistrement {  
2     type zone1;  
3     type zone2;  
4     ...  
5     type zoneN;  
6 };  
7 struct nom_enregistrement variable = {val_zone1, val_zone2, ..., val_zoneN}
```

Exemple:

```
1 ▸ struct etudiant {  
2     int matricule;  
3     char nom[50];  
4     char prenom[50];  
5     float moyenne;  
6 };  
7 struct etudiant e1 = {123456, 'Nour', 'Ahmed', 15.75}
```



Fichier (1 / 2)

- Ensemble organisé d'informations, désigné par un nom précis, que le système d'exploitation d'un ordinateur manipule comme une simple entité, dans sa mémoire centrale ou sur une mémoire secondaire.
- Un fichier est un ensemble d'informations homogènes, qui ont une relation logique.
- Un fichier est un ensemble de champs (zone) regroupés sous forme d'enregistrements.
- Un fichier peut être conservé de façon permanente dans une mémoire secondaire.
- Exemples:
 - ▣ Un fichier employeur contient les informations des employés d'une entreprise.
 - ▣ Un fichier texte est une séquence de caractères organisés en différentes lignes
 - ▣ Un fichier image est une séquence de pixels.
 - ▣ Un fichier code-source est une séquence de sous-programmes.
 - ▣ Un fichier objet est un ensemble des mots organisés en blocs.

Fichier (2/2)

Fichier = Métadonnées + Contenu

- Métadonnées: paramètres utilisés par le système d'exploitation:
 - **Nom:** identificateur utilisé pour distinguer différents fichiers dans la machine par un être humain.
 - **ID:** identificateur utilisé pour distinguer différents fichiers dans la machine par le système d'exploitation.
 - **Type:** séquence de caractères décrivant nature de contenu (texte, image, son, vidéo).
 - **Chemin:** pointeur sur la position où le fichier est enregistré.
 - **Volume:** la taille de fichier en terme d'octets.
 - **Protection:** mode d'autorisation d'accès au fichier (lecture seule, lecteur écriture)
 - **Date/Temps:** date et temps de création, ou dernière modification du fichier.
 - **ID utilisateur:** identificateur de créateur du fichier.
 - **Autres:** données spécifiques aux type de fichiers: longueur/largeur d'une image etc.

Fichier/Enregistrement logique

- Le fichier logique est décrit par sa structure (la structure des enregistrements qu'il contient).
- Le fichier logique ne dépend pas du support physique qui va être utilisé pour le stockage du fichier.
- Les enregistrements d'un fichier logique sont dits enregistrements logiques ou articles.
- A chaque article est associée une clé qui l'identifie de façon unique. C'est ce qu'on appelle la clé primaire.
- La taille d'un enregistrement est mesurée en octets ou en caractères; elle peut être fixe, variable ou indéfinie.
 - ▣ **Taille fixe:** fichier d'employés
 - ▣ **Taille variable:** commentaires
 - ▣ **Taille indéfinie:** fichier texte

Fichier/Enregistrement physique

- Le fichier physique est le résultat de stockage d'un fichier logique sur un support physique.
- Un fichier physique est un ensemble de blocs où un bloc est constitué d'un certain nombre d'enregistrement.
- Les enregistrements d'un fichier physique sont dits enregistrements physiques.
- Un bloc représente l'unité de transfert entre la RAM et la mémoire secondaire.
- Pour l'exploitation du fichier, les métadonnées d'un fichier sont rangées généralement dans un bloc spécial du fichier appelé bloc d'en-tête.
 - ▣ **Informations statiques:** nom du fichier, date de création, etc.
 - ▣ **Informations dynamique:** nombre courant d'enregistrements, l'adresse du dernier bloc, etc.

Fichier physique vs. Fichier logique en C

```
1 FILE *f;  
2 f = fopen ("C:/.../fichier.txt", "w");
```

Nom logique de
fichier

Connection entre le fichier
logique et le fichier
physique

Le fichier physique

Fichier statique vs. Fichier dynamique

- **Fichier statique:** un fichier qui subit très peu d'insertions et de suppressions.
- **Fichier dynamique:** un fichier où les insertions et les suppressions sont très fréquentes.

Fichier mono-volume vs. Fichier multi-volume

□ Fichier mono-volume :

- un fichier qui est stocké sur un seul support de stockage ou volume.
- toutes les données du fichier sont contenues dans un seul emplacement, qu'il s'agisse d'un disque dur, d'une clé USB, d'un CD-ROM, etc.
- concerne les petits fichiers ou lorsque les données tiennent facilement sur un seul support.

□ Fichier multi-volume :

- un fichier qui est divisé en plusieurs parties distinctes, chacune étant stockée sur un support de stockage différent.
- chaque volume est une partie du fichier, mais le fichier complet est reconstitué en combinant tous les volumes.
- souvent utilisée pour gérer des fichiers volumineux qui ne peuvent pas tenir sur un seul support.

Volume multi-fichier

- **Volume multi-fichier :**
 - ▣ ensemble de fichiers distincts, mais étroitement liés, qui sont généralement organisés pour être gérés ensemble.
 - ▣ peuvent contenir des parties d'un ensemble de données plus large, par exemple, une base de données peut être divisée en plusieurs fichiers, chacun contenant une table différente.
 - ▣ peuvent être utilisés pour améliorer la gestion des données et permettre une répartition logique des informations.



Caractéristiques d'un fichier

Caractéristiques d'un fichier

- **Taille (Volume):** la quantité d'information (nombre de caractères) contenu dans ce fichier exprimée en Octet.
- **Mesures d'activités:** l'ensemble des mesures sur les différentes manipulations effectuées sur le fichier :
 - ▣ **Taux de consultation:** nombre d'enregistrement consultées divisé par nombre total d'enregistrement.
 - ▣ **Fréquence de consultation:** nombre d'accès à un enregistrement du fichier pour simple consultation ou mise à jour.
 - ▣ **Taux de renouvellement :** nombre relatif de nouveaux enregistrements qui sont insérées dans le fichier.
 - ▣ **Stabilité :** un fichier est dit stable pendant une période si le nombre d'enregistrement créés est approximativement égal au nombre d'enregistrements supprimées.

Caractéristiques d'un fichier

- **Accroissement:** nombre d'enregistrement créés par rapport à celui des enregistrements supprimés.
- **Facteur de blocage:** taille d'enregistrement logique divisée par taille d'enregistrement physiques. Il définit le nombre d'enregistrements logique dans un enregistrement physique (bloc physique).

Caractéristiques d'un fichier: Exercice

- Considérons les zones suivantes: *Numéro de sécurité social, Nom, Prénom, Date de naissance, Adresse, Fonction.*
 1. Quel est le sujet commun de ces zones ?
 2. Donnez un code, une taille et un type à chaque zone ?
 3. Donnez le facteur de blocage de ce fichier si un bloc physique est de 100 octets ?
 4. A chaque mois une liste des employés est envoyée au service paie pour la préparation des salaires ; donner la fréquence de consultation de ce fichier ?
 5. Pour l'année en cours trois changements d'adresse ont été effectués; donner la fréquence de consultation de ce fichier pour cette année ?
 6. Pour 1000 employés calculer le taux de consultation annuelle dans les conditions des questions 4 et 5 ?



Topologie des fichiers

Topologie des fichiers

- **Fichier permanent:** un fichier dans les informations ont une importance vitale au sien d'une application informatique, son contenu ne subit pas de fréquentes modification.
- **Fichier mouvement:** un fichier qui sert à mettre à jour un fichier permanent, à la fin du traitement il n'est pas utile.
- **Fichier manœuvre:** il est crée à chaque fois ou une insuffisance en mémoire centrale est constatée, pour contenir seulement les données nécessaires aux traitements. Sa durée de vie est limitée par celle des traitements cause de sa création.
- **Fichier intermédiaire:** sert à contenir les résultats d'un traitement donné qui sera utilisé durant le même traitement, ou par des traitements ultérieurs (changement des données entre programmes).
- **Fichier archive:** sert à conserver les données concernant plusieurs période de la durée de vie d'un programme.



Opérations de bases sur les fichiers

Opérations de bases sur les fichiers

- ❑ **Création de fichier** : un fichier peut être créé dans un système de fichiers en utilisant un gestionnaire de fichiers ou un logiciel approprié. Lors de la création d'un fichier, vous spécifiez souvent son nom, son emplacement et son extension.
- ❑ **Ouverture de fichier** : permet au système d'exploitation de donner accès au contenu du fichier à une application ou à un utilisateur.
- ❑ **Lecture de fichier** : consiste à extraire des données du fichier pour les utiliser dans un programme ou pour les afficher à l'utilisateur. La manière de lire un fichier dépend du langage de programmation utilisé ou de l'application en question.
- ❑ **Écriture de fichier** : consiste à ajouter ou à mettre à jour des informations dans un fichier. Cela peut inclure l'ajout de nouveaux enregistrements ou la modification de données existantes.

The diagram shows a code snippet for opening a file in C. Callouts identify parts of the code: 'Fichier logique' points to the file pointer declaration, 'Fichier physique' points to the file name, and 'Fonction d'ouverture' points to the `fopen` function. A large callout box explains the file modes.

```
1 FILE *fp;  
2 fp = fopen ("emplacement_fichiler", "mode");
```

Fichier logique

Fichier physique

Fonction d'ouverture

r: lecture seule.
w: écriture seule/création.
a: append (ajout à la fin)/création.
r+: lecture et écriture à la fois, le fichier doit être existé.
w+: lecture et écriture. Le contenu est effacé si le fichier existe.
a+: lecture et écriture à la fin du fichier.

Opérations de bases sur les fichiers

- **Fermeture de fichier** : la fermeture d'un fichier libère les ressources système qui lui sont associées et garantit que les données sont enregistrées correctement. Omettre de fermer un fichier peut entraîner la perte de données ou des problèmes de performance.

```
3 fclose(fp);
```

- **Suppression de fichier** : retire généralement définitivement du système de fichiers, bien que des méthodes de récupération puissent être utilisées pour restaurer un fichier supprimé.

```
5 remove("emplacement_fichier");
```

- **Déplacement et renommage de fichier** : déplacer un fichier d'un emplacement à un autre ou le renommer pour changer son nom.

```
7 rename("ancien_nom", "nouveau_nom");
```

- **Copie de fichier** : permet de créer une copie identique d'un fichier existant.
- **Compression et décompression** : compresser un fichier permet de réduire sa taille, ce qui est utile pour économiser de l'espace disque ou pour faciliter le transfert. La décompression permet de restaurer le fichier à son état d'origine.

Opérations de bases sur les fichiers

- **Éclatement:** consiste à diviser un fichier en plusieurs fichiers distincts en fonction de critères prédéfinis, généralement en utilisant un ou plusieurs champs comme base de séparation.
- **Tri :** une opération qui réorganise les données d'un fichier en fonction d'un ou plusieurs critères, généralement en ordre croissant ou décroissant.
- **Fusion:** consiste à combiner deux fichiers ou plus en un seul fichier en utilisant un ou plusieurs critères de correspondance pour aligner les données des fichiers sources.
- **Extraction:** consiste à extraire une partie spécifique des données d'un fichier en fonction de critères de sélection. Cela permet de créer un nouveau fichier contenant uniquement les données répondant aux critères spécifiés.