

Architecture des Ordinateurs et Systèmes d'Exploitation

Cours n°9

Les Fichiers



Ph. Leray



Architecture des Systèmes d'Information

3ème année

Fichiers : pour quoi faire ?

- **La mémoire principale est**
 - de taille restreinte
 - volatile, ...

⇒ **Besoin de stocker des informations :**

- à plus long terme
- de grande taille
- « partageable »

⇒ **Gestion de Fichiers = Partie du système d'exploitation (*file system*)**

- de quelle manière nommer les fichiers,
- comment les structurer,
- comment les utiliser, les protéger ...

Caractéristiques des fichiers

- **Nom : variable selon les SE**
 - lettres de l'alphabet + chiffres et caractères spéciaux
 - longueur
 - différence majuscule / minuscule
 - nom en 2 parties (nom.extension) obligatoire / non
- **Structure :**
 - suite d'octets : taille variable, grande souplesse d'utilisation [Unix|DOS]
 - suite d'enregistrements : enregistrement de taille fixe, moins souple
 - arbre d'enregistrements de taille variable
- **Accès : (dépendant de la technologie du support)**
 - séquentiel : anciens SE
 - accès aléatoire

Types de fichiers

- **Type de fichiers :**
 - **fichiers ordinaires : informations des utilisateurs (fichiers classiques)**
 - » **Fichiers ASCII :**
 - lignes de texte
 - caractère « fin de ligne » variable selon les SE
 - avantage = facilement lisible et modifiable
 - » **Fichiers binaires**
 - structurés pour que l'on puisse retrouver ses informations
 - **catalogues (*directories*) : fichiers systèmes contenant la structure du système de fichiers**
 - **fichiers spéciaux caractères : modélisation d'E/S (terminaux, imprimante,...)**
 - **fichiers spéciaux bloc : modélisation des disques**

Attributs de fichiers

- **Description de base d'un fichier = nom + données**
- **Attribut = information complémentaires :**
 - **liste variable selon le SE**
 - **Ex :**
 - **Protection**
 - **Mot de passe du fichier**
 - **Créateur**
 - **Propriétaire**
 - **Indicateur lecture seule**
 - **Indicateur fichier caché**
 - **Indicateur fichier système**
 - **Indicateur d'archivage**
 - **Indicateur ASCII/Binaire**
 - **Indicateur fichier temporaire**
 - **Indicateur de verrouillage**
 - **Longueur d'enregistrement**
 - **Date de création**
 - **Date du dernier accès**
 - **Date de modification**
 - **Taille courante**
 - **Taille maximale**
 - **...**

Opérations sur les fichiers

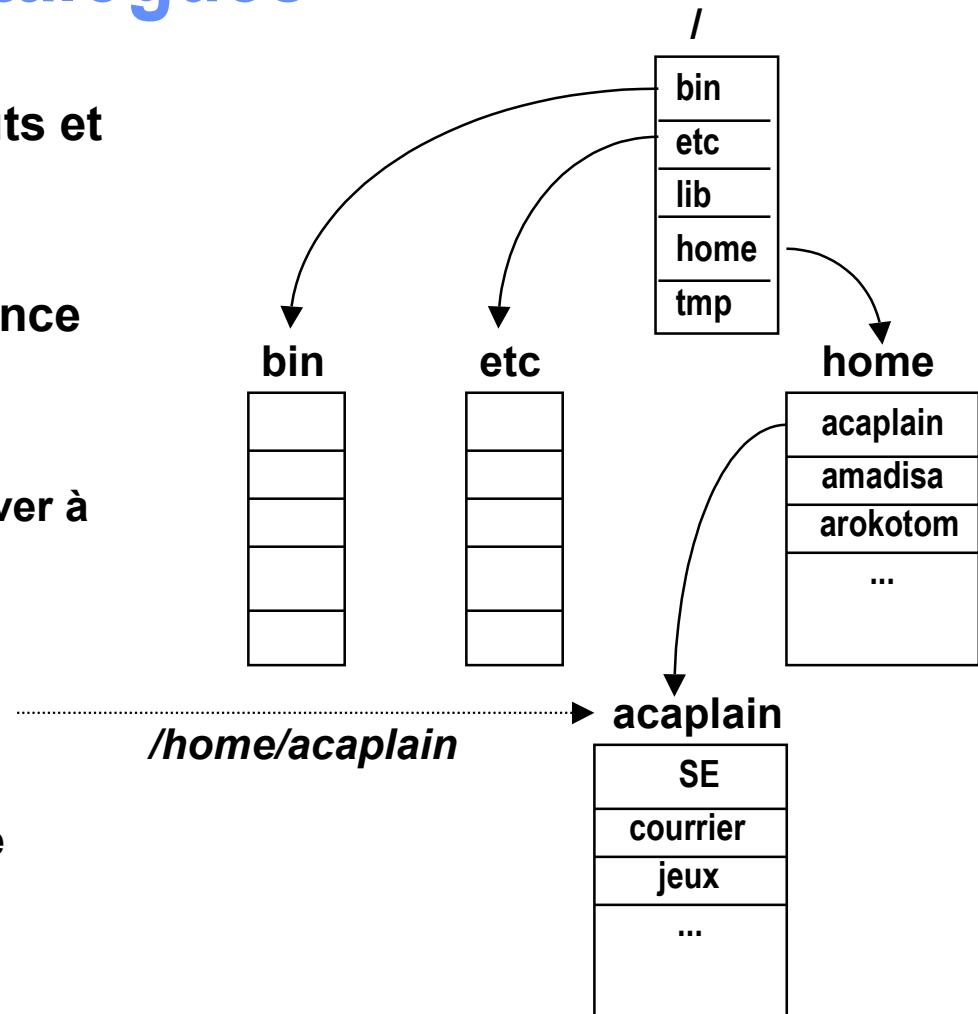
- Ces opérations se font grâce à des appels systèmes :
 - **CREATE** (création d'un fichier vide)
 - **DELETE**
 - **OPEN**
 - **CLOSE**
 - **READ** (lecture de n octets à partir de la position courante)
 - **WRITE**
 - **APPEND** (écriture en fin de fichier)
 - **SEEK** (indication de la position des données à lire/écrire)
 - **GET ATTRIBUTES**
 - **SET ATTRIBUTES**
 - **RENAME**

Catalogues

= Mémorisation des noms, attributs et adresses des fichiers

- Structure courante : arborescence de catalogues
- Chemin d'accès : comment arriver à un fichier situé dans cet arbre ?

- Chemin absolu = à partir de la racine de l'arbre
- Chemin relatif = à partir de la position courante dans l'arbre
- dans certains SE :
 - . = catalogue courant
 - .. = catalogue père



Ex. d'arborescence de catalogues sous Unix

Stockage des fichiers

= Comment le SE sauve le fichier sur le disque

- **allocation contiguë** : tous les blocs se suivent
 - méthode la plus simple mais coûteuse en espace disque
- **allocation «liste chaînée»** : chaque bloc contient l'adresse du bloc suivant
 - pas d'espace perdu mais accès aléatoire lent
- **allocation «liste chaînée indexée»** : regroupement des «adresses de bloc» dans une table (index) en mémoire [MS-DOS]
 - inconvénient = table entièrement en mémoire
- **«nœuds d'information» (*inode*)**: table associée à chaque fichier, contenant les attributs et les adresses des blocs sur le disque [UNIX]
 - avantages = la table est de petite dimension
 - l'i-node contient aussi l'adresse de la liste des blocs vides (utile en cas d'ajout)

Fichiers «partagés»

- Un fichier a souvent besoin d'être dans plusieurs répertoires (A et B) sans être dupliqué à chaque fois
- Avantage du fichier partagé = gain de place
- Problème = si le fichier est changé dans A, comment le catalogue B va-t-il s'en apercevoir ?
 - Pas de problème si les infos fichiers sont dans un i-node
 - Autre solution = lien symbolique
 - » Création d'un fichier de type lien dans B qui contient le chemin d'accès du fichier de A
- Inconvénient = un parcours automatique du disque passera plusieurs fois par les mêmes endroits

Fiabilité d'un système de fichiers

Qui n'a jamais perdu un fichier lors d'un plantage, panne de courant ou autre ?

- **Gestion des blocs endommagés = mémoriser la liste des blocs endommagés pour éviter de les réutiliser !**
 - Matérielle (directement sur une piste du disque)
 - Logicielle (dans un fichier)
- **Sauvegarde**
 - Méthodes « double disque »
 - Sauvegardes incrémentales
- **Cohérence du système de fichier : si une panne survient au moment où un bloc est écrit, le système de fichier devient incohérent**
 - Le SE doit posséder un utilitaire de vérification (et réparation) de la cohérence du système de fichiers

Sécurité et fichiers : pourquoi ?

- **Perte de données :**
 - feu, inondations, ...
 - erreurs matérielles / logicielles
 - erreurs humaines
- **Intrusions humaines :**
 - indiscretion occasionnelle, jeu
 - chantage économique, piratage commercial, militaire
- **Droit :**
 - protection de l'utilisateur contre un mauvais usage des informations le concernant
 - interdiction de « croiser » des fichiers

Sécurité et fichiers : comment ?

- **Contre les pertes d'informations :**
 - sauvegarde des données dans un lieu différent
 - logiciels de récupération de données sur les disques
 - «commandes simples» :
 - Erreur de frappe classique (taper `rm p *` au lieu de `rm p*`)
 - Pour l'éviter : alias `rm = rm -i`
- **Contre les intrusions humaines :**
 - logiciels anti-virus sur des systèmes peu protégés (Windows)
 - identification de l'utilisateur + accès restreint à ses fichiers et commandes
 - contre-mesures = un compte est bloqué lorsque 3 essais de connexion ont échoué, ...
 - tests de sécurité par des experts
- **Contre l'utilisation des informations :**
 - CNIL

Références

Gestion de fichiers : généralités

- Architecture de l'Ordinateur - A. Tanenbaum (InterEditions)

Gestion de fichiers : UNIX/Linux :

- UNIX : Utilisation, Administration, Réseau Internet - C. Pélissier (Hermès)
- UNIX : Guide de l'étudiant - H. Hahn (Dunod)
- Le Système Linux - M. Welsh (O'Reilly)
- Linux pour les nuls - C. Witherspoon et al. (Sybex)

Sécurité - Administration Unix:

- Les bases de l'administration système - A. Frisch (O'Reilly)

A suivre :

- TP n° 9a : Comparaison Gestion de Fichiers Windows/Linux
- TP n° 9b : Les fichiers sous Unix

Les supports hardware

Comment stocker de l'information ?

Petit rappel

Avantages

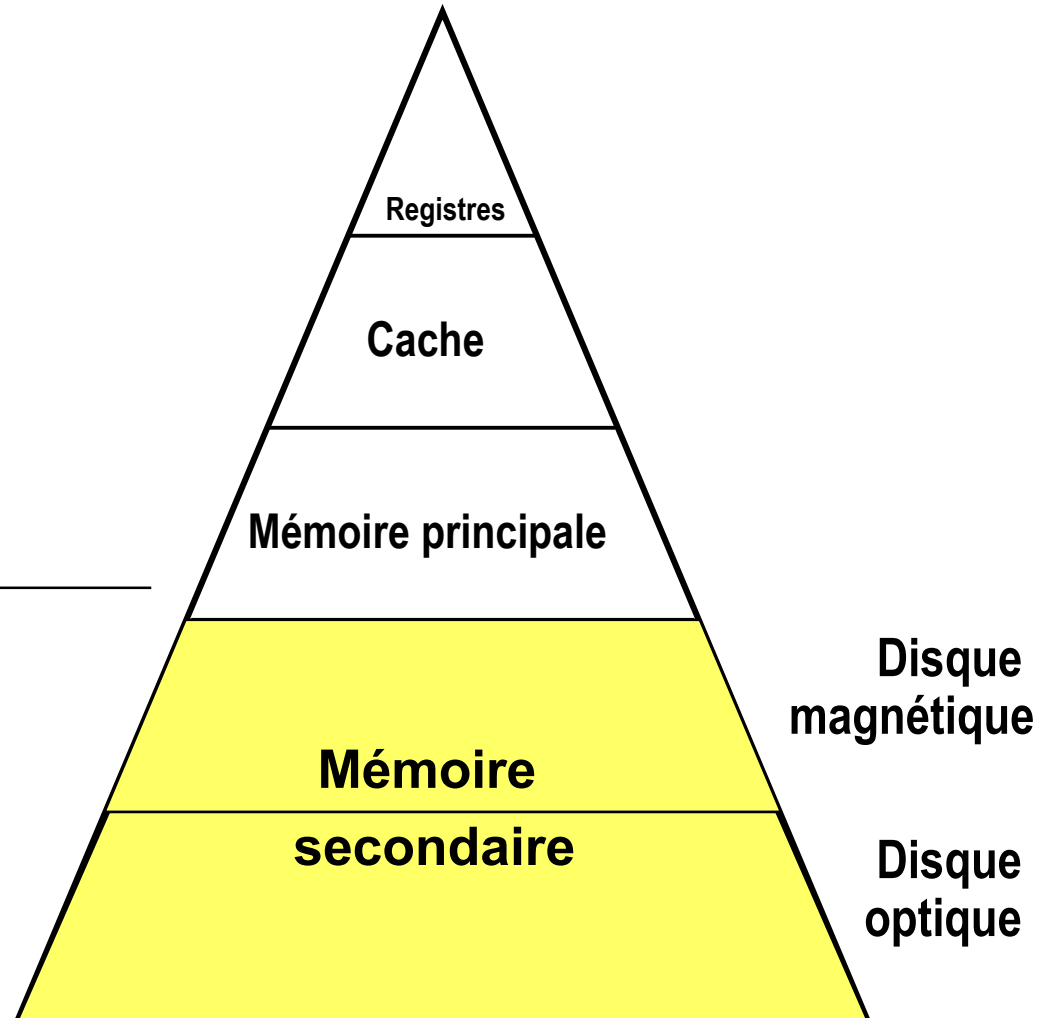
Rapidité
Accès direct
Faible taille

Inconvénients

Volatile
Faible capacité
Chère (prix/bit)

Peu chère
Non volatile
Grande capacité

Lenteur
Volumineuse
Accès séquentiel / mixte



Plan

- **Typologie des mémoires secondaires**
- **Comparatif de quelques supports**
- **La disquette**
- **Le disque dur**
- **La technologie RAID**
- **Du CD au DVD**

Typologie des mémoires secondaires

- **Technologie :**
 - Optique | Magnétique

- **Support :**
 - Disque | Bande
 - Fixe | Amovible

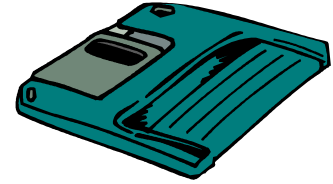
- **Caractéristiques :**
 - Capacité | Prix /Go
 - Temps d'accès
 - Temps de transfert
 - Temps moyen de fonctionnement sans panne

Magnétique		Optique
Bande	Disque	Disque
DAT	3,5 "	Réinscriptible
Exabytes	5,25 "	WORM
Cartouches	RAID	CD

Comparatif de quelques supports

Support Technologie	Capacité (Go)	Prix/Go	Taccès	Ttransfert
Disquette (3,5 ")	.001	2000 F		
Zip 100 (disque amovible)	0.1	800 F	29 ms	1.4 Mo/s
Jazz (disque amovible)	1	600 F	12 ms	7 Mo/s
Syquest (disque amovible)	5	300 F	12 ms	10 Mo/s
Exabyte (bande magnétique)	7	15 F		
Disque dur (3,5 ")	20	75 F	9 ms	200 Mo/s
DAT (bande magnétique)	24			1 Mo/s

La disquette



- **Format courant : 3"1/2**
- **Support magnétique + enveloppe rigide**
- **Mode d'enregistrement :**
 - Simple densité (SD) : dépassé
 - Double densité (DD) : 720 Ko
 - Haute Densité (HD) : 1,44 Mo
 - Extra Haute Densité (ED) : 2,88 Mo
- **Indétrônable ?**

Le disque dur

- Empilement de disques magnétiques 3"1/2 ou 5"1/4
- Ex (assez) récents :

Marque	Taccès	Tr/min	Prix HT
Fujitsu 20.4 Go EIDE	10 ms	5400	1800
Quantum Katana 18.2 Go EIDE	9.5 ms	7200	2000
Fujitsu 18.2 Go SCSI	8 ms	7200	4000

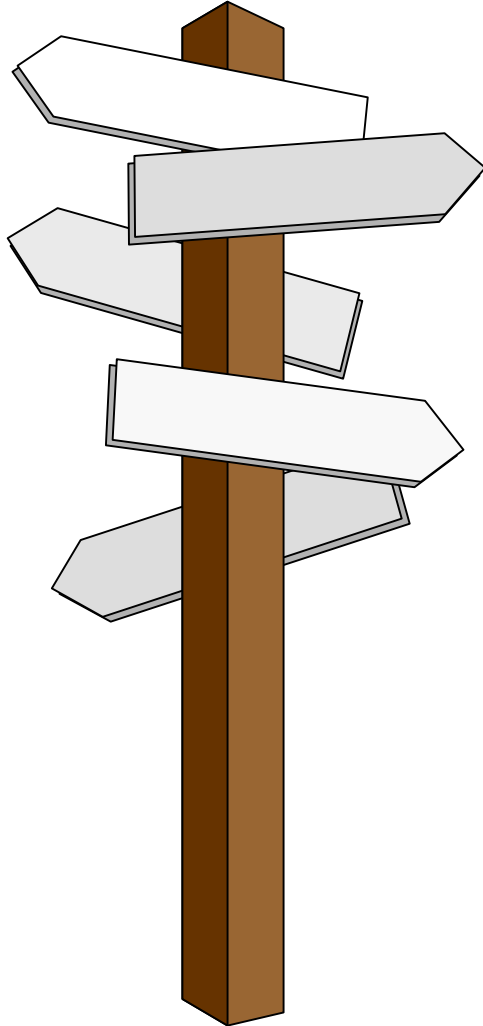
La technologie RAID

- **RAID = Redundant Array of Inexpensive (Independant) Disks**
- **Principe = utilisation de plusieurs disques pour sauvegarder les informations et les récupérer facilement**
- **Plusieurs niveaux de sécurité (1 à 6)**
 - miroir
 - contrôle d 'erreur
 - découpage de l 'écriture sur plusieurs disques, ...

Du CD au DVD

- **Support = disque optique numérique**
- **CD / CDRom / WORM / CDR ...**
- **Ttransfert dépend du lecteur**
 - **CD 8x = 1200 Ko/s**
 - **CD 32x = 4800 Ko/s ...**
- **(Future) norme = DVD**
 - **DVD Vidéo, DVD-Rom**
 - **DVD-R, DVD Audio, DVD-RW**
 - **Comme tout support, son développement va dépendre du prix du lecteur**

Bibliographie



- **Livres de référence :**
 - Architecture et technologie des ordinateurs - P. Zanella & Y. Ligier (Dunod)
 - Technologie des ordinateurs et des réseaux - P.A. Goupille (Dunod)
- **Sites WWW :**
 - Multimedia Information Sources : <http://dutiem.twi.tudelft.nl/projects/MultimedialInfo/>
 - Le Club Technique PC