

Architecture des ordinateurs

27 - Stockage de masse en réseau

Philippe Darche
IUT Paris Descartes

Besoins (en vrac !)

- Flux croissant d'informations
 - le « zéro-papier »
 - le commerce électronique (*e-business*)
 - etc.
- ⇒ Quantité d'informations croissante à mémoriser
- Partage de l'information interne/externe à l'entreprise
- Disponibilité 24h/24 7j/7
 - tolérance aux fautes
- Sécurité de l'accès
- Gestion centralisée et aisée

Problèmes

- Hétérogénéité des environnements informatiques
 - SE différents
 - les Unix, les Windows, OS/400, pour systèmes embarqués, etc.
 - systèmes de fichiers différents
 - extx, NTFS, HPFS, etc.
 - support de mémorisation aux caractéristiques différentes
 - formats de données différents
 - etc.

Type de stockage (en informatique)

□ Sauvegarde (*backup*)

- duplication et mise en sécurité des données pour une restauration éventuelle en cas de problème

⇒ copie temporaire

- opération symétrique : restauration

□ Archivage

- procédé de conservation des données fiable et fidèle dans le temps

⇒ copie permanente

- opération symétrique : récupération

Type de sauvegarde

- Totale
 - sauvegarde de tous les fichiers
- Différentielle
 - sauvegarde de tous les fichiers créés ou modifiés depuis la dernière sauvegarde totale
- Incrémentale
 - sauvegarde de tous les fichiers créés ou modifiés depuis la dernière sauvegarde



Sauvegardes

- Locale
 - sauvegarde de production
- Externalisée
 - sauvegarde de recours

De quoi parle-t-on ?

- Baie de disques magnétiques
- Systèmes à bandes magnétiques
 - *Striped Tape*
 - *Automated Media Handling*
 - *MSS (Mass Storage System) : Automated Tape Library*
- Unité de mémoire optique

Baie indépendante de disques

- *Disk arrays*
- Assemblage de disques offrant des fonctionnalités :
 - augmentées : débit, capacité, etc.
 - supplémentaires : tolérance aux fautes, surveillance, etc.
- Interface d'E/S rapide
- Séparée du ou des CPU
 - séparation fonctionnelle calcul/stockage
 - séparation machine utilisateur/serveur applicatif/mémorisation de masse
- Mise en œuvre de techniques de détection et de correction d'erreurs
 - ajout d'informations de contrôle \Rightarrow perte de capacité utile
 - perte de bande passante si réparation
- Gestion séparée
 - sauvegarde
 - surveillance
 - réparation
- Alimentation électrique indépendante

Mais attention

- Fiabilité (*reliability*)

- MTBF (*Mean Time Between Failures*) en heure

- ordre de grandeur (fourchette basse) : 500 000 h

- fiabilité de d disques = fiabilité d'un disque / d

- ⇒ nécessité obligatoire de mécanisme de détection/correction d'erreurs

Plusieurs approches de tolérance aux fautes

- Miroir/recopie (*Mirroring/Shadowing*)
 - coût de stockage élevé
- Utilisation des codes détecteur et correcteur d'erreurs
 - basés sur la parité
 - détection
 - codes de Hamming
 - correction
- Service de prévision de fautes
 - surveillance (*monitoring*)
 - pas de perte de capacité
 - ancêtre : VaxSimPlus (*VAX System Integrity Monitor*)
 - technologie actuelle : SMART (*Self-Monitoring Analysis and Reporting Technology*)
 - issue de la technologie PFA (*Predictive Failure Analysis*) d'IBM
 - attributs de l'unité surveillés en interne, avertissement du SE via le BIOS en cas de dépassement de seuil(s)

Autre solution

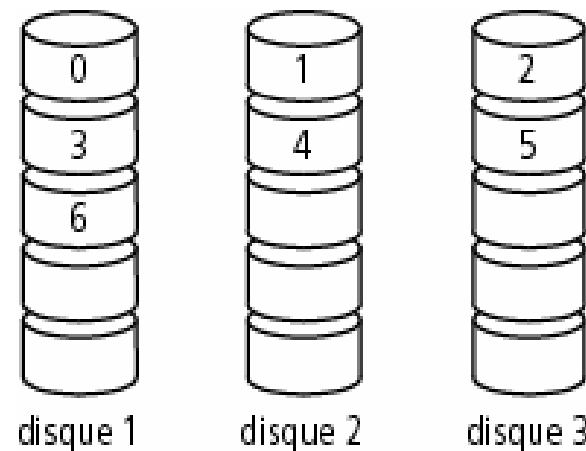
- Disque de secours (*hot spare*)
avec reconstruction en parallèle
 - solution de luxe mais très haute disponibilité

RAID

- *Redundant Arrays of Inexpensive Disks*
 - Articles de référence : [Patterson et al. 88]
- Rédéfinition du terme par [Gibson 92]
 - *Redundant Array of Independent Disks*
- Permet de voir un ensemble de disques physiques comme un seul disque logique
- Numérotation de 0 à 5 (6 niveaux)
 - combinaison possible
 - exemple : RAID 0-1
- Gestion matérielle ou logicielle

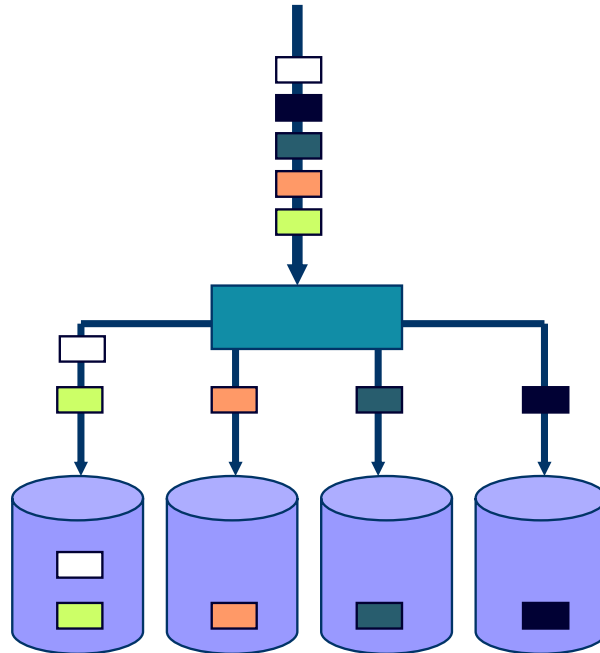
RAID 0

- Entrelacement de disques
ou volume agrégé par bandes (*data striping*)
- Données réparties sur d disques
 - accès en parallèle
 - ⇒ bande passante augmentée d'un facteur d
 - volume maximal
 - mais aucune redondance des informations
 - ⇒ pas de tolérance aux fautes !



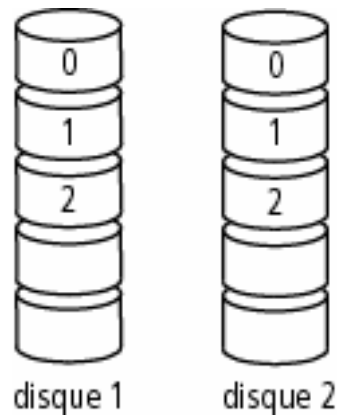
RAID 0

Exemple RAID 0



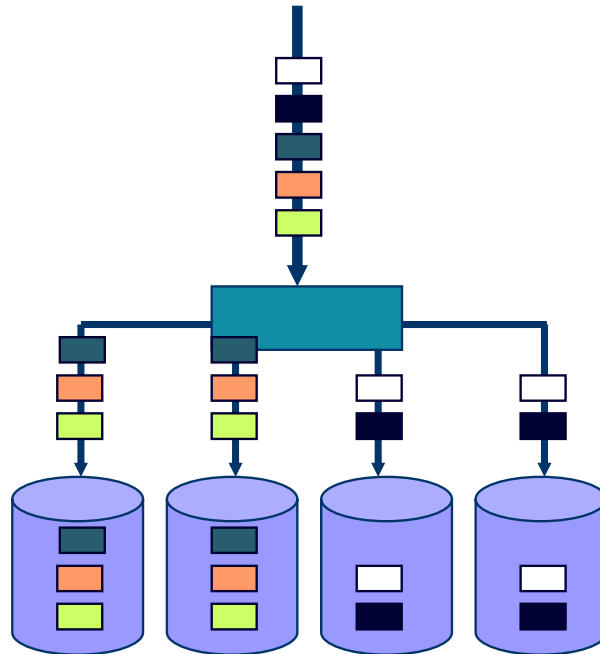
RAID 1

- Disques en miroir (*Disk mirroring or shadowing*)
 - une écriture logique = deux écritures physiques



RAID 1

Exemple RAID 1



RAID 1

□ Avantages

- solution la plus rapide avec tolérance de panne
- taux de duplication maximal
- insertion à chaud et remplacement à chaud possibles

□ Inconvénients

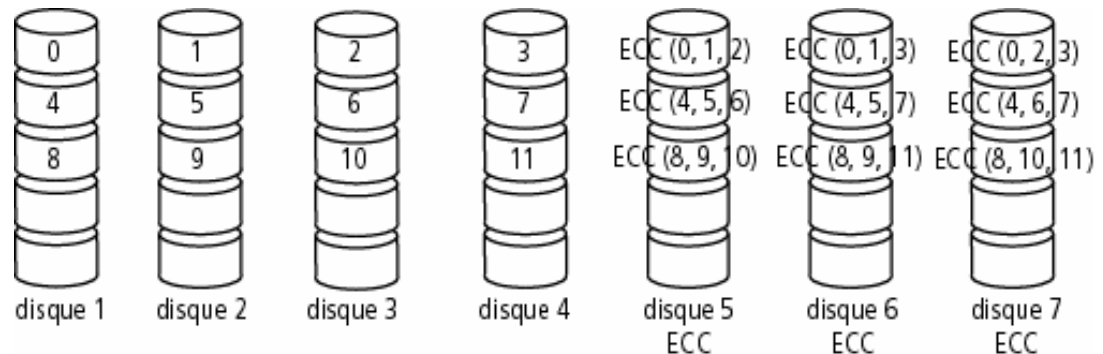
- division par 2 de l'espace disque
- pas d'accélération de l'écriture comme en RAID 0

RAID 2

- ❑ RAID 0 et correction d'erreur
- ❑ Utilisation de disques dépourvus de système de correction d'erreur (code de Hamming)

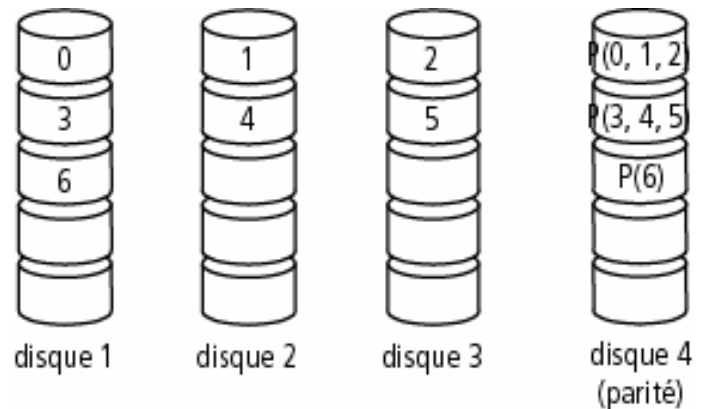
- ❑ **Obsolète**

- coût élevé
- matériel spécialisé

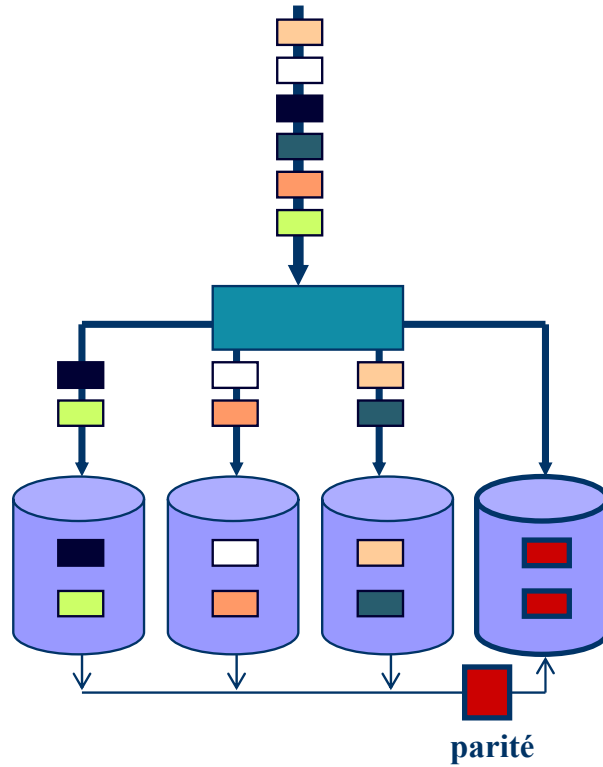


RAID 3

- Répartition des données « en bandes »
comme en RAID 0
- Un disque de parité
 - goulot d'étranglement
- entrelacement
au niveau de l'octet



Exemple RAID 3 ou 4

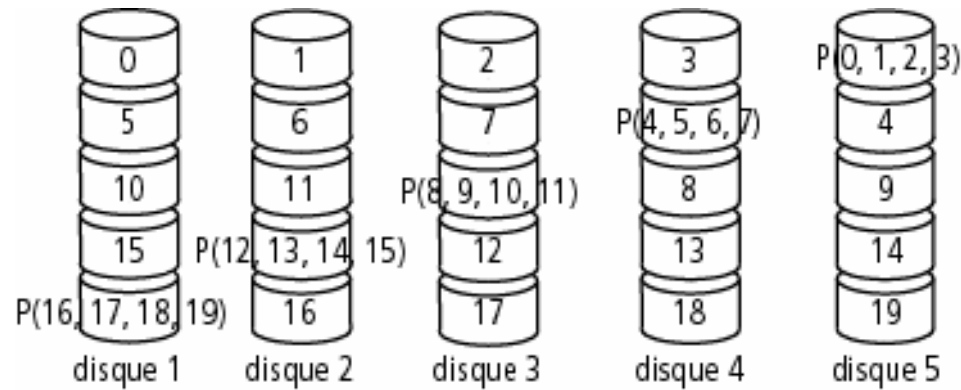


RAID 4

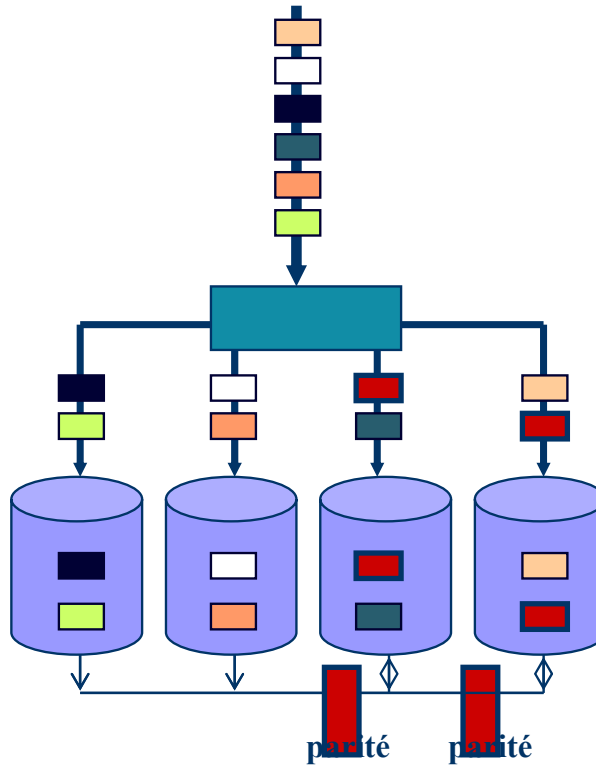
- Même principe que le RAID 3
- Mais entrelacement au niveau du bloc (*Block Interleaved Parity*)
- Lectures/écritures indépendantes (asynchrone)

RAID 5

- Similaire au RAID 4 mais répartition de la parité sur les autres disques (*Block-Interleaved Distributed Parity*)
- Reconstruction difficile
 - amélioration : disque *spare*



Exemple RAID 5



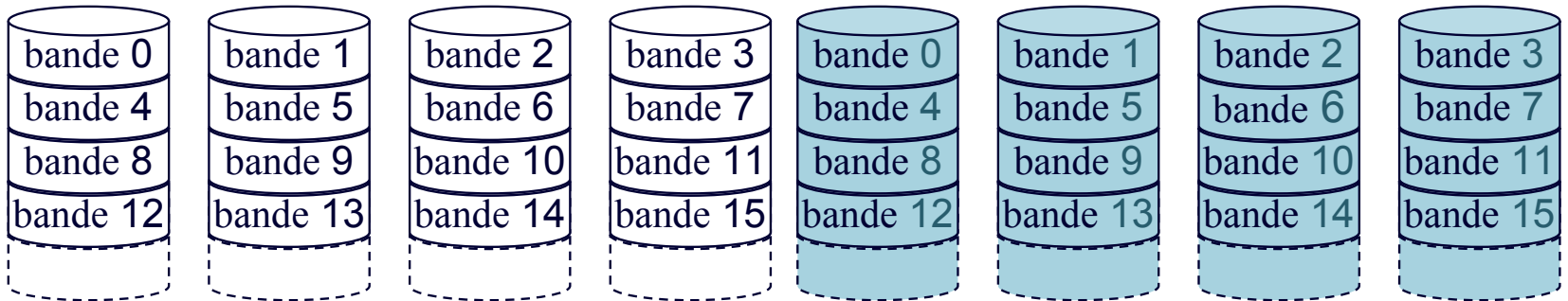
RAID6

- Utilisation du code reed-solomon
- $P + Q$ Redundancy

Association possible

- Bandes en miroir (*mirrored stripes*)

- RAID 0+1 ou 01



- Miroirs en bande (*striped mirrors*)

- RAID 1+0 ou 10

- 0+5, 5+0 (*striped parity*), 1+5, 5+1, etc.



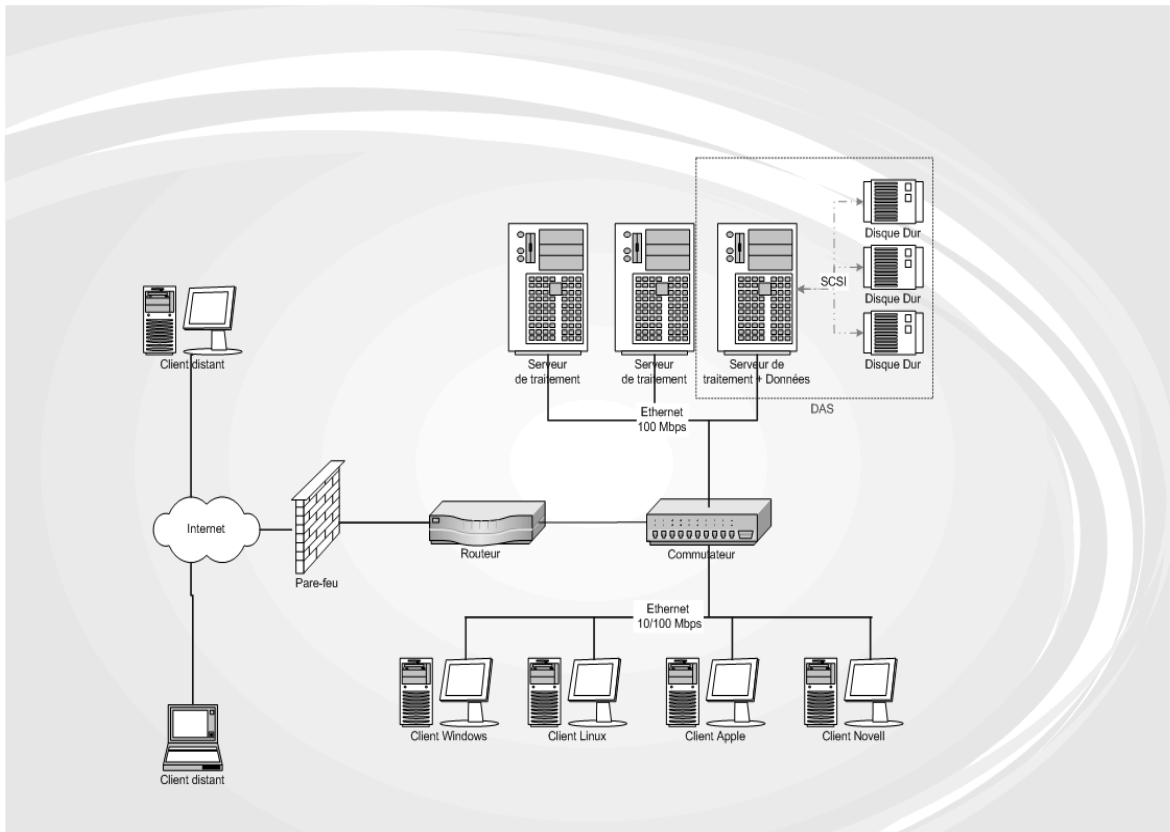
Les trois approches de stockage

- DAS
- NAS
- SAN

DAS

- *Direct Attached Storage (devices)*
- Système de stockage directement raccordé à un serveur
 - interface classique (PCI/SATA/SCSI)
 - bande passante limitée
 - gestion locale
 - bas coût
 - mais problème de passage à l'échelle
- ⇒ adapté pour la petite société ou le particulier

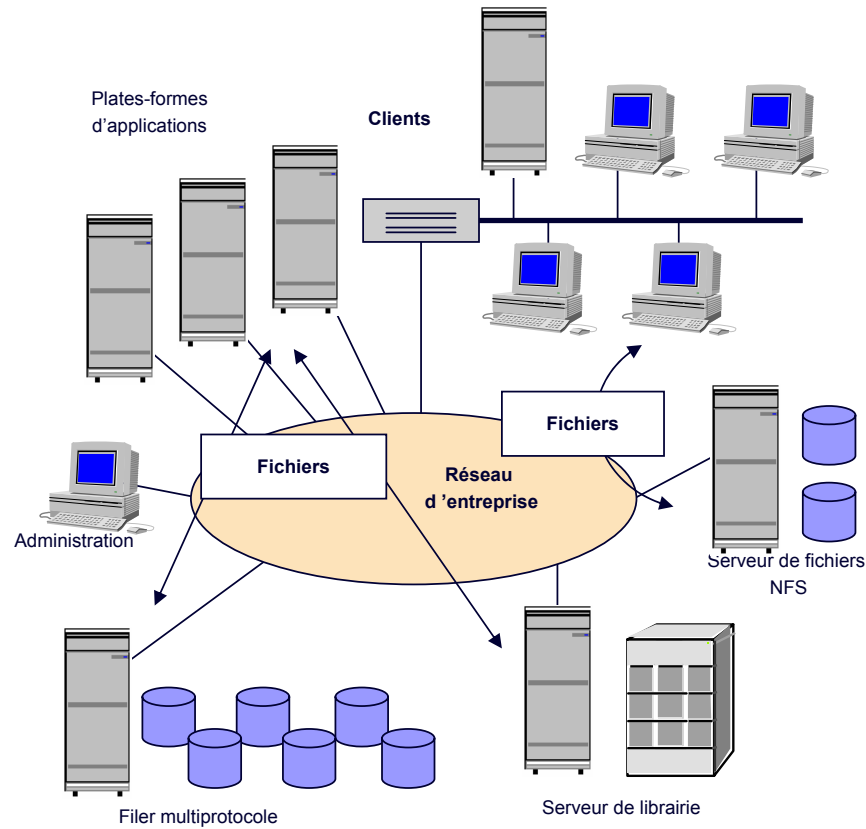
Architecture



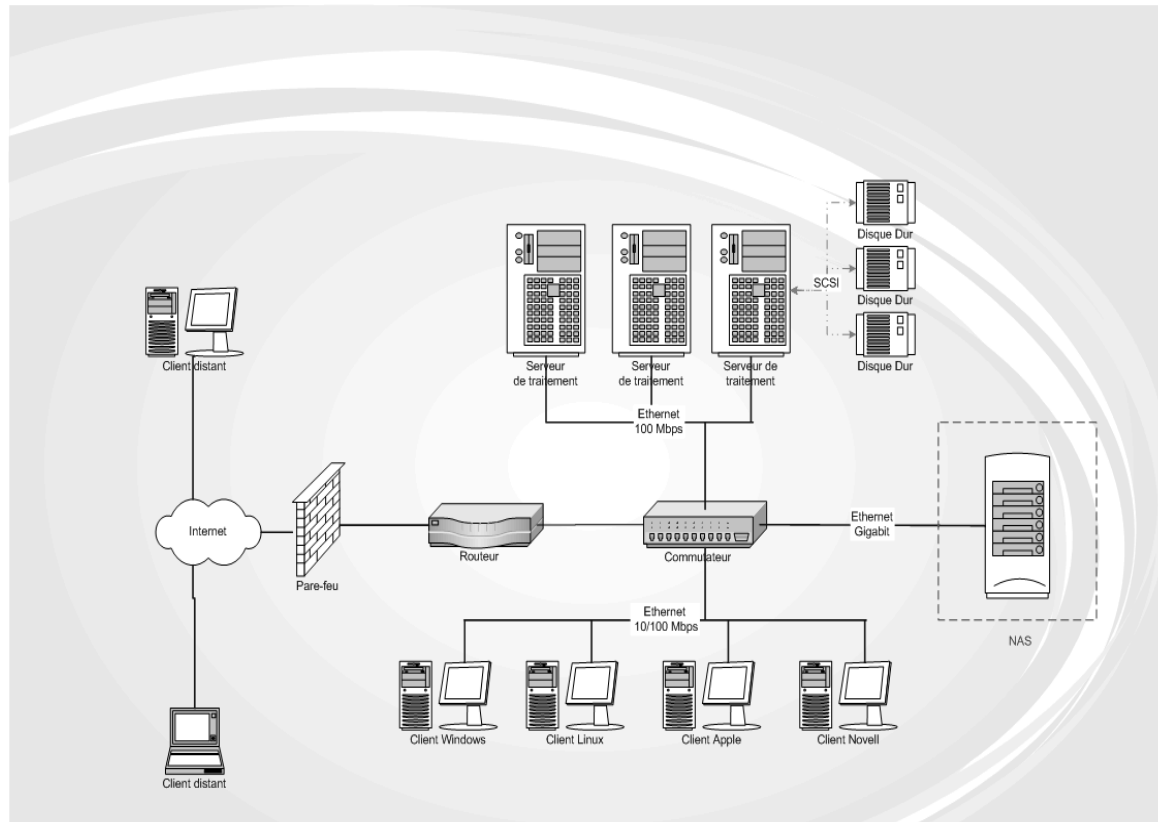
NAS

- *Network Attached Storage*
- Réseau de disques relié par une interface très rapide
- **Services d'accès distants aux fichiers** en natif
comme :
 - NFS (*Network File System*)
 - SMB/CIFS (*Server Message Block/Common Internet File System*)

NAS



Architecture d'un NAS





Intérêts d'une zone de stockage

- Gestion centralisée des données
- Détection/correction d'erreurs
- Sauvegarde centralisée

NAS

□ Avantages

- coût faible
- facilité d'installation et d'administration
- connectable à un réseau pré-existant
- grande capacité de stockage de 500 Go à 10 To
- disque à insertion à chaud (*hot plug*)
- assure l'intégrité des données et la redondance du système
- idéal pour les réseaux hétérogènes

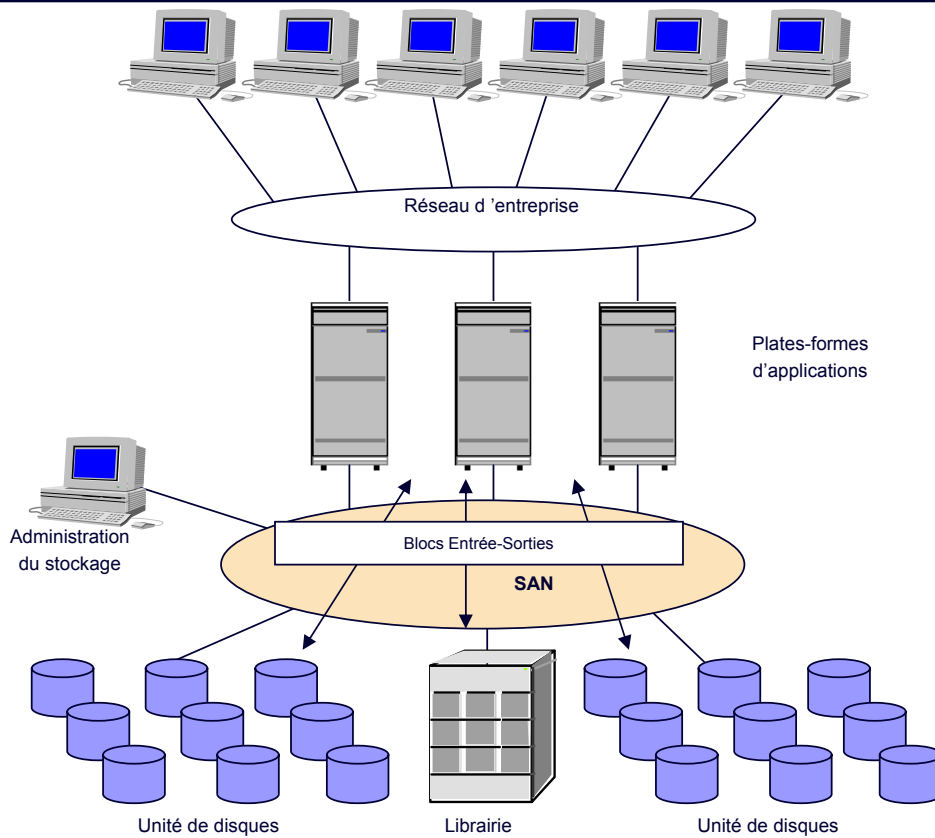
□ Inconvénient

- induit un risque de surcharge et d'encombrement du réseau

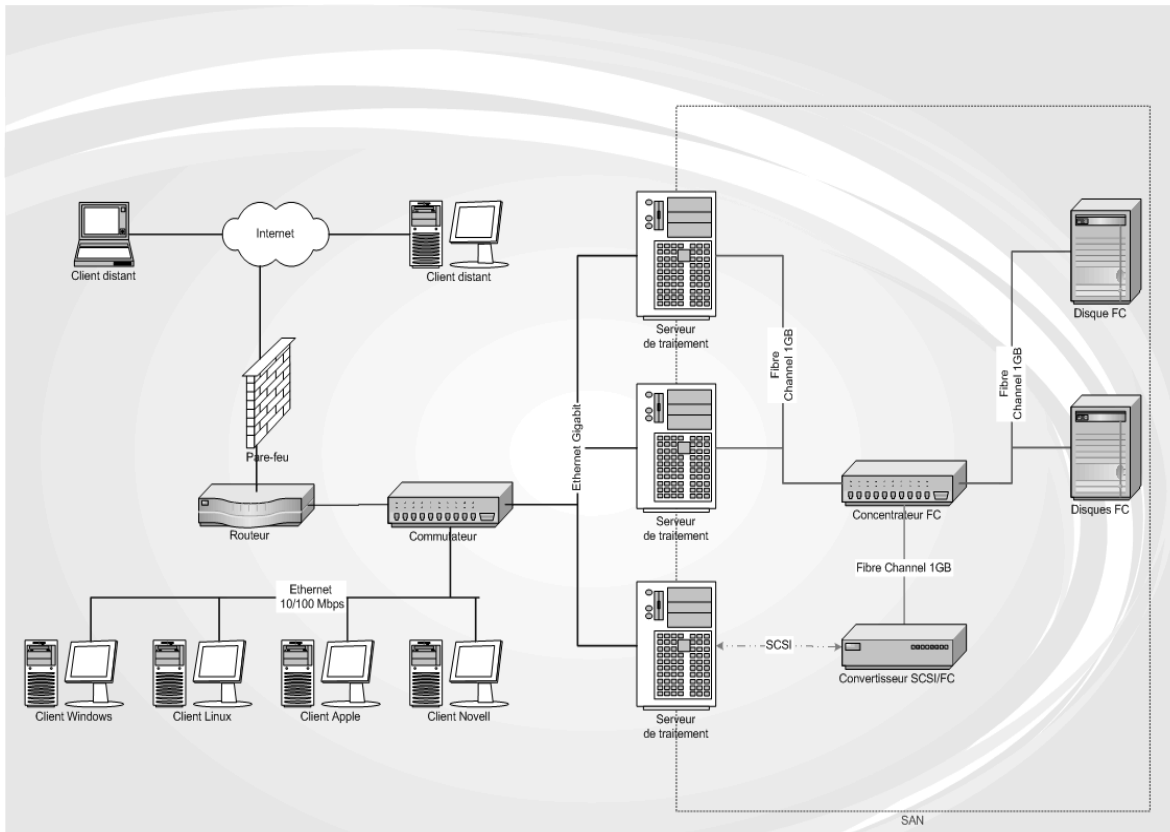
SAN

- *Storage Area Network*
- Système de stockage (lecteurs de bande et baies de disques RAID)
 - connecté à des serveurs via un réseau à très haut débit qui fournit un service de type bloc
- Accès aux fichiers (*file-level access*) grâce au système de fichiers SAN (*SAN filesystem* ou *shared disk filesystem*)

SAN



Architecture d'un SAN



Connexions possibles

- Fibre Channel
 - commutateur (*switch* ou *director*)
 - le plus répandu
- iSCSI (*internet Small Computer System Interface*)
 - protocole SCSI encapsulé dans des trames TCP/IP
- FCoE (*Fibre Channel over Ethernet*)
 - protocole Fibre Channel via ethernet

Arbitrated Loop

- FC-AL (*Fibre Channel Arbitrated Loop*)
- Anneau unidirectionnel à 127 nœuds maximum
 - réalisation physique avec un HUB

SAN

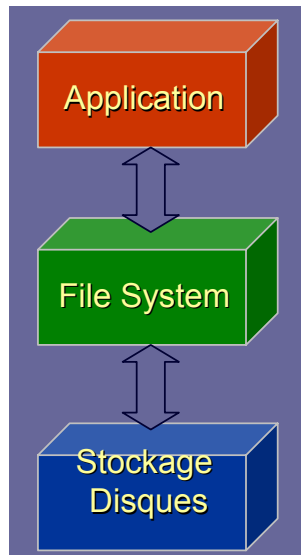
□ Avantages

- solution performante
- haute disponibilité des données (intégrité + redondance)
- LAN déchargé de l'accès aux données

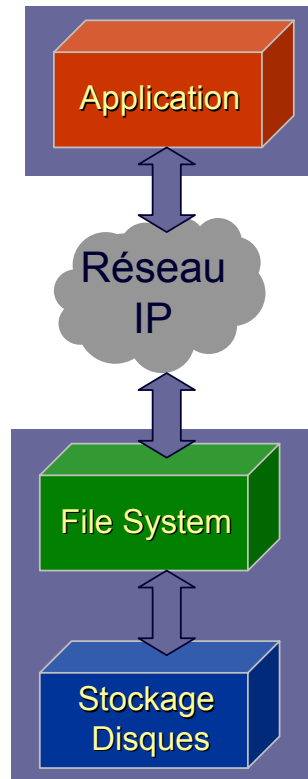
□ Inconvénients

- déploiement souvent complexe
 - câblage, répartiteur, etc.
- prix excessivement élevé

Synthèse : Les trois approches possibles

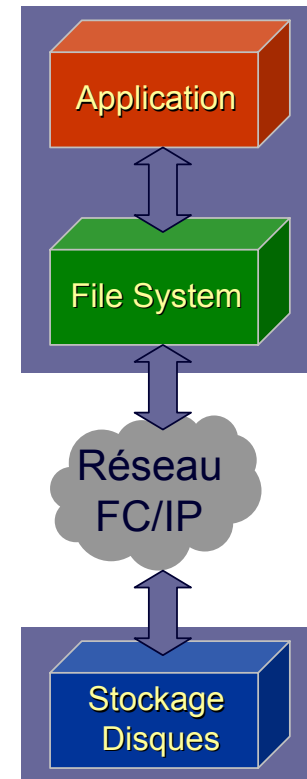


Serveur



Client

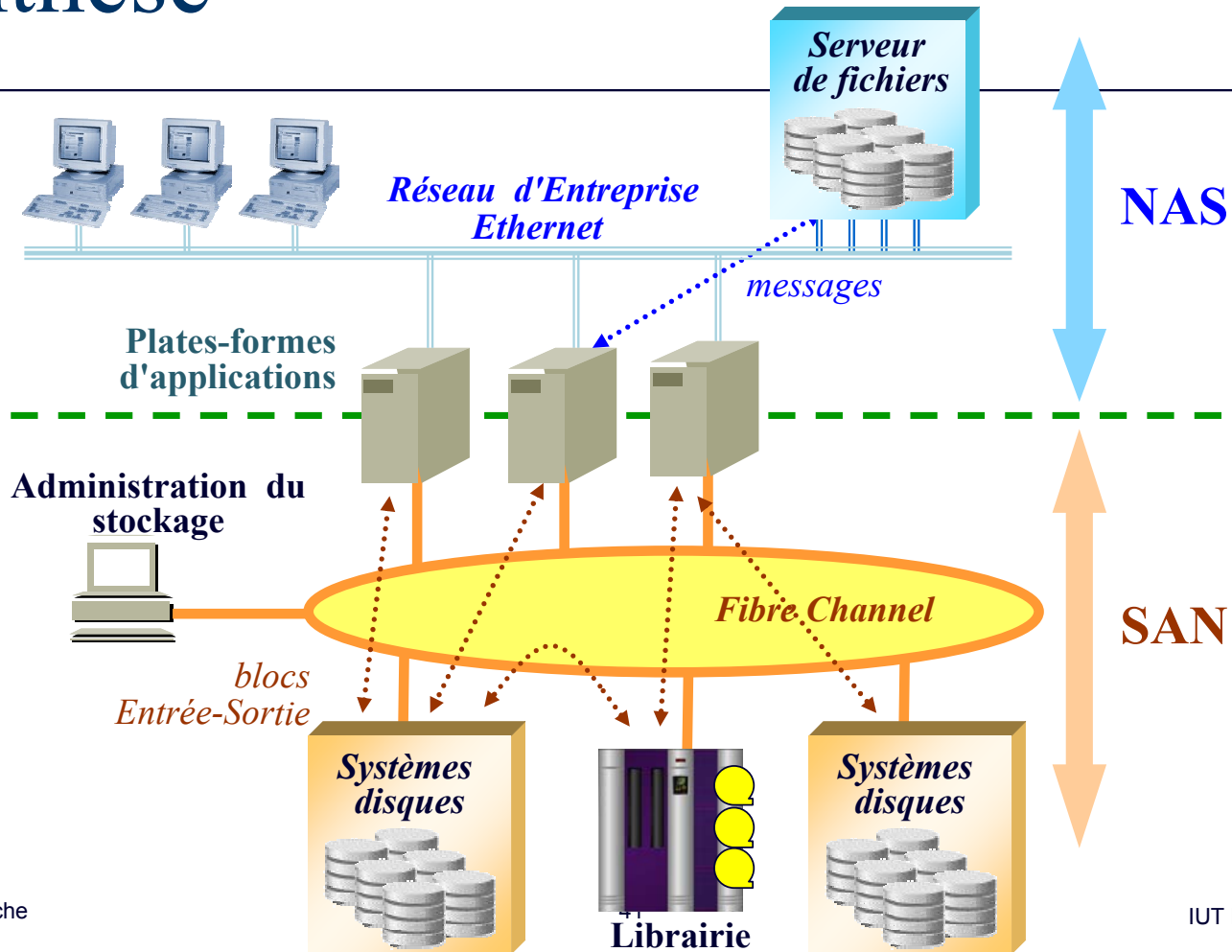
Serveur NAS



Serveur Application

Sous système
RAID

Synthèse



Conclusion

- La hiérarchie mémoire est modifiée

