Réseaux

Sauvegarde

Sommaire

Notions

La réalité de la sauvegarde :

- 58% des PME n‘ont pas de stratégie de reprise d'activité après sinistre ;
- 50% des entreprises ont déposé le bilan dans les 36 mois après un sinistre ;

Les menaces :

- Inondation ;
- Séisme ;
- Panne électrique ;
- Incendie ;
- Grève ;
- Pannes matérielles ;

Business Impact Analysis (BIA) :

- Perte financière ;
- Perte de marchés ;
- Perte de crédibilité ;
- Perte de l‘image de marque ;
- Défaillance des prestataires ;

Identification de la reprise d‘activité :

- Personnes ;
- Compétences ;
- Services nécessaires ;
- Délais de redémarrage minimum ;
- Délais de redémarrage complet ;

Contraintes de production :

- Le SI fonctionne 24 / 7 (cf. cours sur la haute-disponibilité) ;
- Les volumétries sont croissantes ;
- Le temps d'indisponibilité est réduit ;
- La réglementation oblige à rendre des comptes (Sarbanes-Oxley, Bâle II, LSF, etc...) ;

→ Effet positif : implique les décideurs dans le processus de sauvegarde

Stratégies de sauvegarde

Différence entre sauvegarde d'un poste individuel et sauvegarde d'un serveur

Méthodes de sauvegarde différentes pour plusieurs raisons :

- les données sur **poste client** sont réputées **moins importantes** que les données gérées sur des systèmes centraux ;
- les utilisateurs sont moins sensibilisés au risque de perte de données que les professionnels de l'informatique ;
- les moyens techniques sont moins développés sur poste individuel que sur serveur.

La sauvegarde des données des postes individuels reste marginale dans la stratégie d'utilisation des ordinateurs.

Inscription dans une démarche plus globale → continuité d'activité

Formalisation dans un document → PRA, PCA, plan de secours, ...

PCA : Plan de Continuité d'Activité

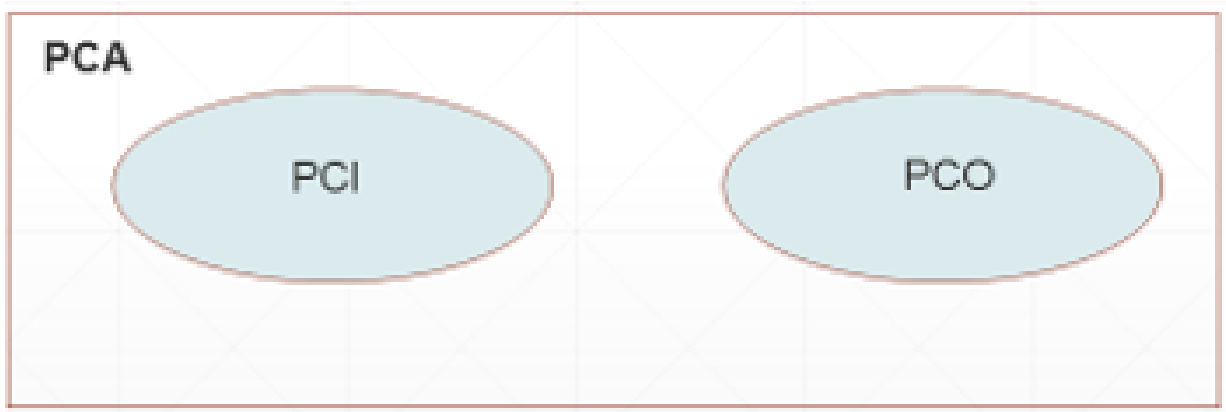
Permet, en cas de crise, de continuer l'activité sans perte de service ou avec une dégradation acceptable

PRA : Plan de Reprise d'Activité

Permet, en cas de sinistre, de pouvoir reconstruire ou basculer sur un système de secours pour une durée déterminée dans le but d'assurer la survie de l'entreprise

Le PCA intègre:

- le PCI : Plan de Continuité Informatique
- le PCO : Plan de Continuité des Opérations (orienté métier)



RTO:

Le Recovery Time Objective est le temps maximal acceptable pendant lequel une ressource informatique peut ne pas être fonctionnelle :

RPO:

Le Recovery Point Objective désigne la durée maximum d‘enregistrement des données qu‘il est acceptable de perdre lors d‘une panne.

Différents types de sauvegarde

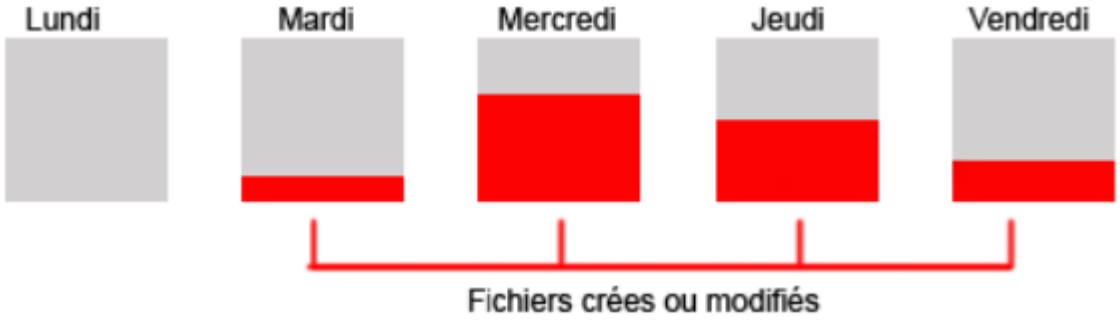
Sauvegarde totale :

- Sauvegarde toutes les données, les répertoires et les sous répertoires sélectionnés.
- La plus rapide, simple et précise pour **restaurer** les données sans erreurs.
- Prend beaucoup de temps à faire → souvent trop long
- Peu géner l‘activité de l‘entreprise

Différents types de sauvegarde

Sauvegarde incrémentale :

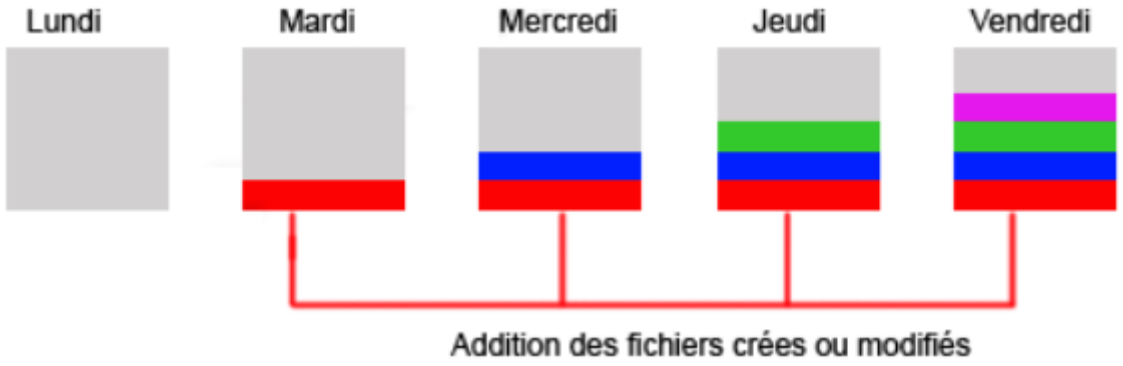
- sauvegarde uniquement les données qui ont été modifiées ou ajoutées depuis la dernière sauvegarde
- volume de données sauvegardées plus faible
- temps d‘exécution plus court
- nécessite moins d‘espace de stockage
- complexe et longue à restaurer



Différents types de sauvegarde

Sauvegarde différentielle :

- sauvegarde les données qui ont été modifiées ou ajoutées depuis la dernière sauvegarde
- prend plus de temps qu‘une sauvegarde incrémentale
- n‘offre pas de rémanence (dernier état d‘un fichier uniquement)
- restauration très rapide



Archivage :

- Conserver les données sauvegardées
- Procédures encadrées par la loi et les normes → non répudiation
- Capitalisation de l‘activité → réutilisation dans des projets futurs
- Respect de la réglementation (CNIL)

Espaces de sauvegarde :

- Local, directement sur le poste
- Sur disque externe : modulable, performant, redondance possible

Mode de sauvegarde :

- Fichier : dans un dossier sauvegardé, si le fichier est modifié il est sauvegardé
- Bloc : sauvegarde uniquement des clusters modifiés

Le cloud propose différentes solutions :

- Cloud professionnel
- Cloud privé
- Cloud hybride

Le cloud professionnel :

- Sauvegarde depuis un agent local
- Transfert des données au travers d'Internet
- Restauration lente → certains prestataires proposent l'envoi des données par colis
- Confidentialité → oblige le chiffrement des données sauvegardées

Le cloud privé :

- Sauvegarde local sur un NAS / SAN
- Application d'une stratégie de sécurité (anti-X) avant le transfert local ou sur le support de sauvegarde lui-même
- Réfractaires au mutualisé → sauvegarde au domicile (chez le RSSI)

Le cloud hybride :

- Sauvegarde depuis un agent sur un NAS / SAN local
- Application d'une stratégie de sécurité (anti-X) avant le trasfert local ou sur le support de sauvegarde lui-même
- Transfert complet ou partiel sur des équipements présents dans un datacenter
- Restauration rapide car locale
- Si un sinistre survient, récupération possible depuis un site de secours

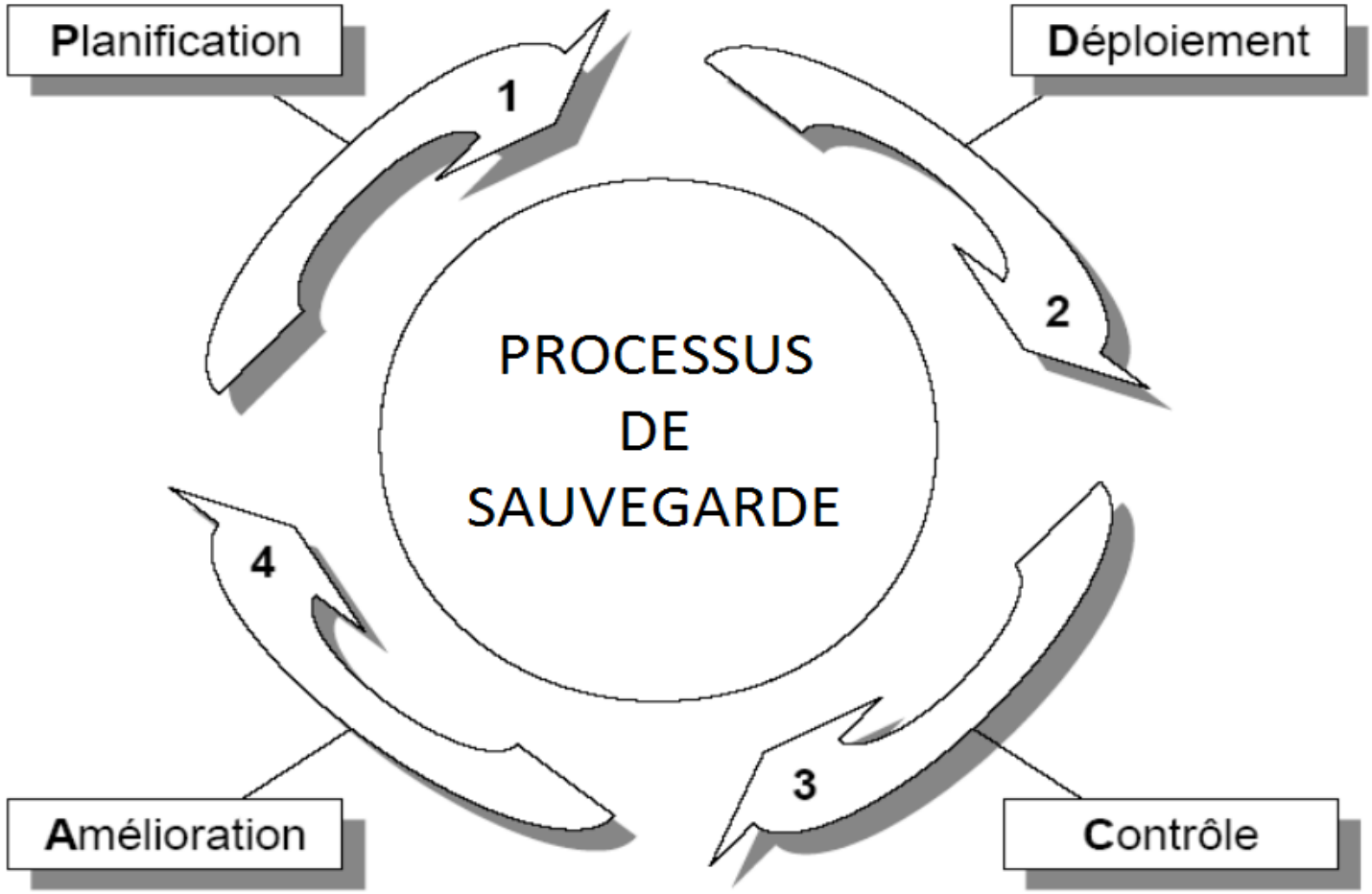
Critères de sélection :

- Volumétrie
- Taux de transfert (**I**nput/output **O**perations **P**er **S**econd)
- Pérénnité (fiabilité du support)
- Archivage (insensibilité aux sinistres)
- Confidentialité (chiffrement)
- Facilité / granularité de restauration
- Contrainte réglementaire (SLA)
- Coût d‘acquisition / exploitation

Système de management :

- Implication de l'ensemble des acteurs
- Travail transversal
- Passage de l'oral à l'écrit
- Processus audité → évaluation constante
- Améliore la qualité de service → respect des bonnes pratiques
- Garantit une certaine qualité (partenaires / clients)

Approche PDCA :

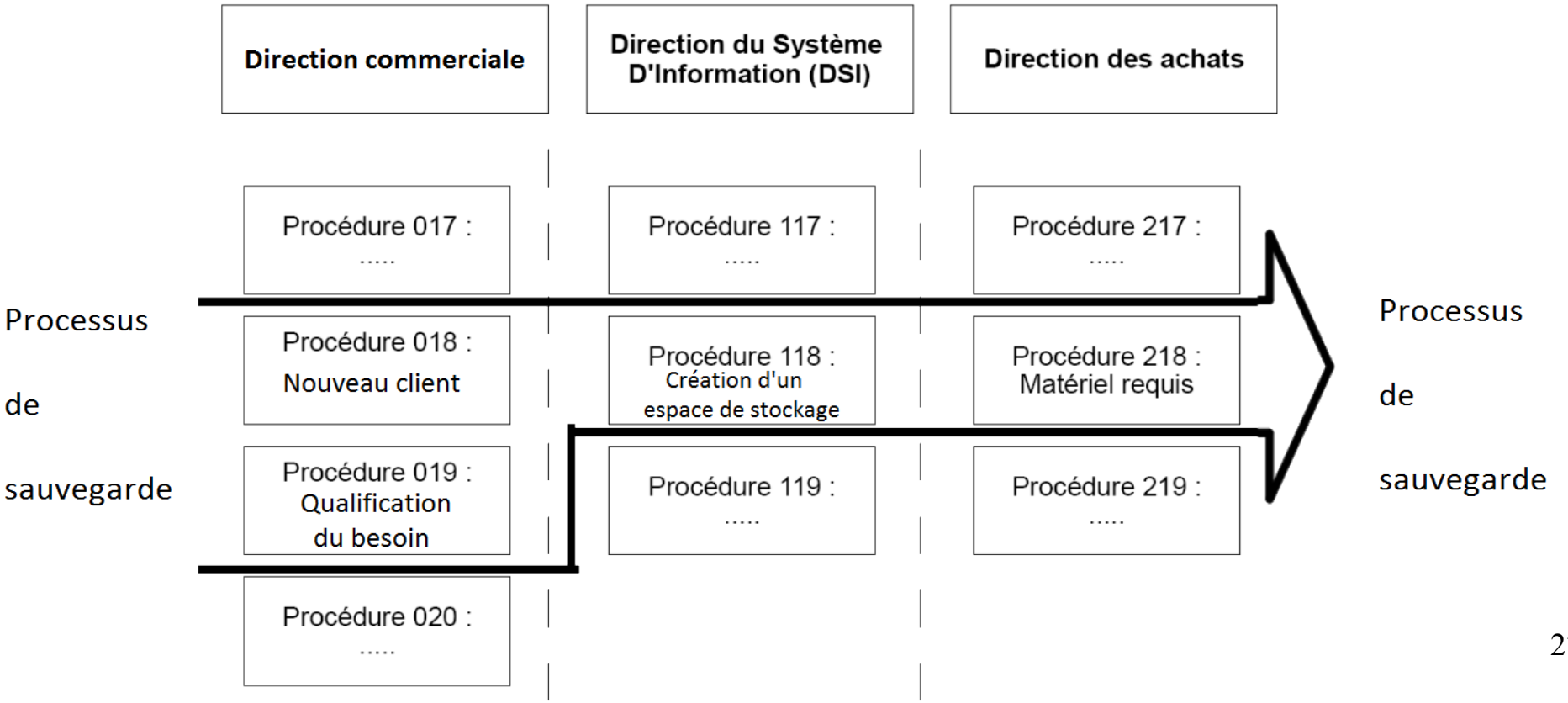


Approche PDCA, différentes phases :

- **Planification** : elle consiste à préparer un programme et un calendrier en fonction des objectifs fixés ;
- **Déploiement** : elle doit donner les moyens financiers, techniques et humains définis dans la phase de planification ;
- **Contrôle** : il vérifie que les objectifs fixés dans la phase de planification sont bien mis en place et, s'il ne le sont pas, à calculer les écarts ;
- **Amélioration** : elle sert à mettre en place les actions correctives qui vont permettre de diminuer les écarts calculés dans la phase précédente.

Insertion dans un "framework" (ISO 27001) :

- Définition des procédures
- Méthode transversale
- Permet de clarifier / optimiser les processus



Mise en place du **S**ystème de **M**anagement de la **Q**ualité (SMQ) :

- Impliquer la direction ;
- Identifier les processus clés ;
- Planification / Documentation du SMQ ;
- Déploiement du SMQ ;
- Contrôle du SMQ ;
- Amélioration du SMQ ;

Solutions techniques

Les disques durs :

Le mécanisme le plus utilisé pour redonder l'information au niveau des disques durs est le RAID (Redundant Array of Independent Disk) en opposition au SLED (Single Large Expensive Disk).

Le RAID permet de mixer des disques durs "classiques" dans une matrice qui sera plus performante (plus d'espace, plus rapide, plus sécurisée)

Il existe deux types de RAID : matériels et logiciels

Les niveaux de RAID qui vont nous intéresser sont 0, 1, 5 et 10

Le RAID logiciel

Dans ce cas, le contrôle du **RAID** est **intégralement assuré** par une **couche logicielle** du système d'exploitation. Cette couche s'intercale entre la couche d'abstraction matérielle (pilote) et la couche du système de fichiers.

Avantages

- **peu cher** ;
- très **souple** (administration logicielle) ;
- la grappe est **compatible** avec toutes les machines utilisant le même OS.

Le RAID logiciel

Inconvénients

- la couche d'abstraction matérielle peut manquer de fonctions importantes comme la détection et le diagnostic des défauts matériels et/ou la prise en charge du remplacement à chaud (Hot-swap) des unités de stockage ;
- la gestion du RAID **monopolise des ressources systèmes** (CPU et bus système)
- l'utilisation du RAID sur le disque système n'est pas toujours possible.

Le RAID logiciel

Implémentations

- Sous Windows XP et + seulement le RAID 0 et 1 sont gérés et sous Windows Serveur le RAID 5 est supporté ;
- Sous MAC seulement le RAID 0 et 1 sont supportés ;
- Sous Linux (noyau 2.6) les RAID 0, 1, 4, 5, 6 et 10 sont supportés ainsi que les combinaisons de ces modes (ex. 0+1)

Le RAID matériel

Une carte ou un composant est dédié à la gestion des opérations, doté d'un processeur spécifique, de mémoire, éventuellement d'une batterie de secours, et est capable de gérer tous les aspects du système de stockage RAID grâce au ***firmware***.

Avantages

- détection des défauts, remplacement à chaud des unités défectueuses, possibilité de reconstruire de manière transparente les disques défaillants ;
- charge système allégée ;
- la vérification de cohérence, les diagnostics et les maintenances sont effectués en arrière-plan par le contrôleur sans solliciter de ressources système.

Le RAID matériel

Inconvénients

- les contrôleurs RAID matériels utilisent chacun leur propre système pour gérer les unités de stockage et donc aucune donnée ne pourra être récupérée si le contrôleur RAID n'est pas exactement le même (firmware compris) ;

- les cartes d'entrée de gamme possèdent des processeurs peu puissants et donc les performances sont moins bonnes ;

- l'entrée de gamme se situe aux alentours de 200€ mais les cartes plus performantes dépassent souvent les 1 000€.

- le contrôleur RAID est lui-même un composant matériel, qui peut tomber en panne ("***single-point-of-failure***") ;

Le RAID 0

Egalement connu sous le nom d'« entrelacement de disques » ou de « volume agrégé par bandes » (***stripping***), c'est une configuration permettant d'augmenter significativement les performances de la grappe en faisant travailler **N** disques durs en parallèle (avec **N > 2**).

- Capacité

La capacité totale est égale à celle du plus petit élément de la grappe multiplié par le nombre d'éléments présents dans la grappe.

L'espace excédentaire des autres éléments de la grappe restera inutilisé, il est donc conseillé d'**utiliser des disques de même capacité**.

Le RAID 0

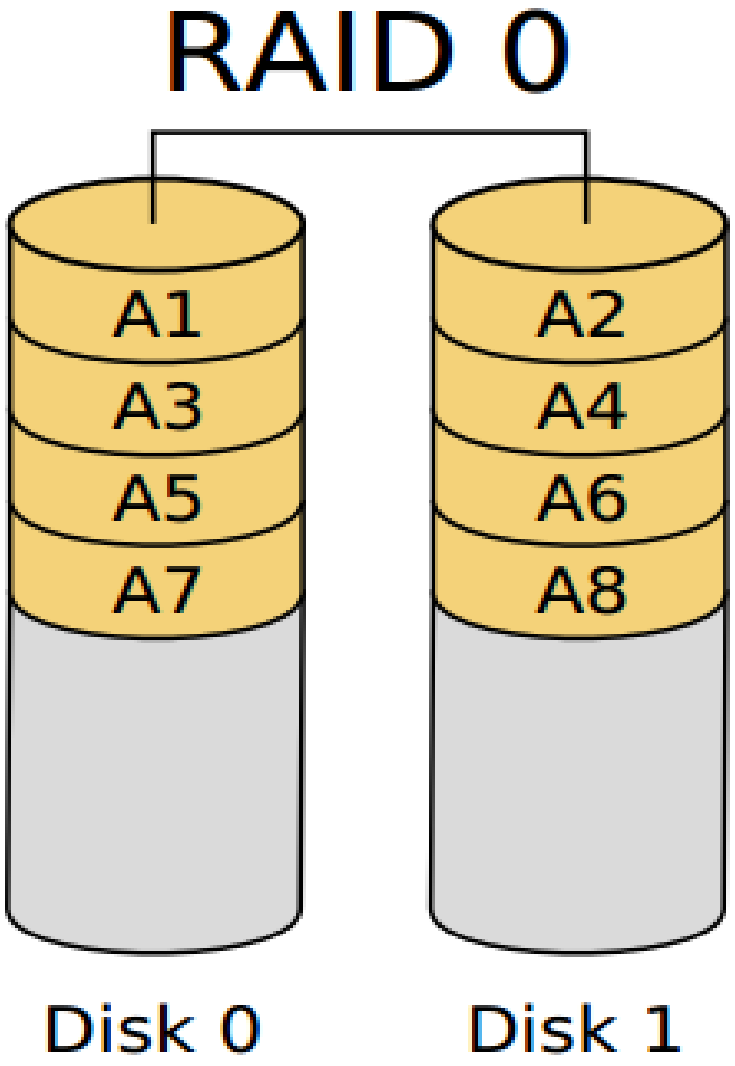
- Fiabilité

Le défaut de cette solution est que la perte d'un seul disque entraîne la perte de toutes ses données.

- Coût

Dans un RAID 0, qui n'apporte aucune redondance, tout l'espace disque disponible est utilisé (tant que tous les disques ont la même capacité).

Le RAID 0



Le RAID 1

RAID miroir (mirroring) consiste en l'utilisation de N disques redondants (avec N > 2), chaque disque de la grappe contenant à tout moment exactement les mêmes données.

- Capacité

La capacité totale est égale à celle du plus petit élément de la grappe. L'espace excédentaire des autres éléments de la grappe restera inutilisé. Il est donc conseillé d'utiliser des éléments identiques.

Le RAID 1

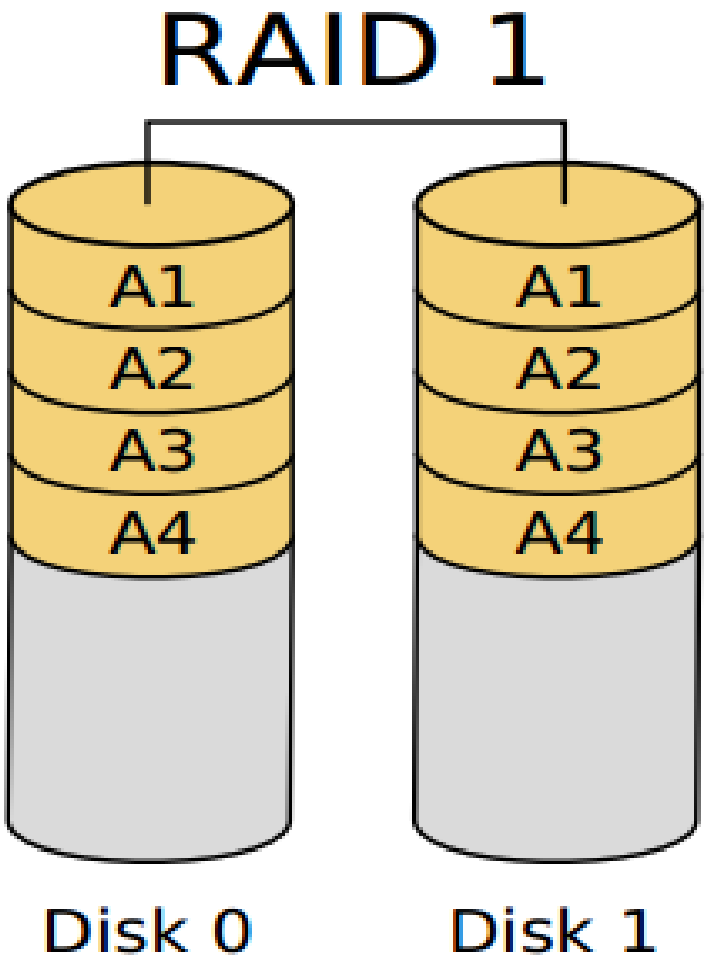
- Fiabilité

Cette solution offre un excellent niveau de protection des données. Elle accepte une défaillance de N-1 éléments.

- Coût

Les coûts de stockage sont élevés et directement proportionnels au nombre de miroirs utilisés alors que la capacité utile reste inchangée. Plus le nombre de miroirs est élevé, et plus la sécurité augmente, mais plus son coût devient prohibitif.

Le RAID 1



Le RAID 5

Le RAID 5 combine le stripping à une parité répartie pour un ensemble à redondance N+1.

La parité, qui est incluse avec chaque écriture, se retrouve répartie circulairement sur les différents disques.

Chaque bande est donc constituée de N blocs de données et d'un bloc de parité.

En cas de défaillance, on peut "retrouver" les données à partir des N-1 autres blocs de données et du bloc de parité.

Pour limiter le risque, il est courant de dédier un disque dit de "spare" qui en régime normal est inutilisé et en cas de panne prendra la place du disque défaillant.

Le RAID 5

- Capacité

La capacité totale est égale à (N-1) x C avec N le nombre de disque et C la capacité du plus petit élément de la grappe.

- Fiabilité

Cette solution offre un excellent niveau de protection des données. Elle accepte une défaillance de N-1éléments.

- Coût

Les coûts de stockage sont corrects et égaux à 1 disque en plus dans le meilleur des cas, à 1 + N disques de spare dans le pire des cas.

Le RAID 5

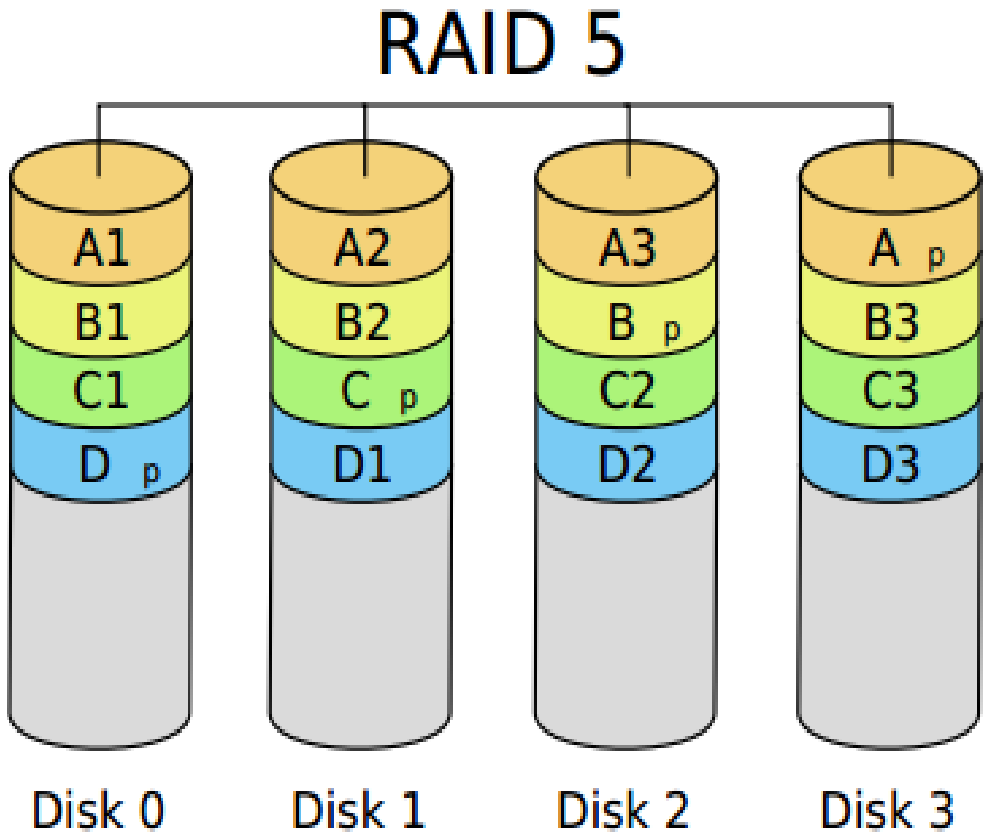

RAID 5
A1
B1
C1
D p
A2
B2
C p
D1
A3
B p
C2
D2
A p
B3
C3
D3
Disk 0
Disk 1
Disk 2
Disk 3

Image d'exploitation :

La virtualisation permet de faire des snapshots (images de machines) virtuels.

Ces snapshots permettent de capturer l'état entier de la machine au moment où ils sont déclenchés.

Ils peuvent être utilisés pour faire un point de contrôle d'un système d'exploitation, sauvant ainsi d'une erreur de configuration.

Copie à chaud :

La copie à chaud permet de sauvegarder la machine virtuelle complète sur un serveur distant, pendant son fonctionnement (aucun downtime)

Gestion de configuration :

La gestion de configuration consiste à gérer la description technique d'un système / équipement

Elle permet la gestion de systèmes complexes ainsi que leurs déploiements.

Exemple :

- rpm -qa ;
- fichiers de configuration des services ;
- configuration des routeurs / switch ;
- …

Gestion itérative grâce à des outils comme SVN

Virtualisation :

Les serveurs de virtualisation n'échappent pas à la régle et sont eux-mêmes sujets aux pannes.

Le cluster ou noeud comprend plusieurs serveurs de virtualisation qui sont dédiés au fonctionnement des **mêmes** machines virtuelles.

Cela permet, à l'image du cloud, de dématérialiser l'endroit où la machine virtuelle s'exécute.

De la sorte, la montée en charge peut être anticipée et la tolérance aux pannes est plus grande.

Cas pratique

Imaginons que l'on veuille restaurer un système en panne :

1) Phase de diagnostic → diagnostiquer le problème et déterminer l'action appropriée.

2) Phase d'approvisionnement → recenser, trouver, transporter et assembler physiquement le matériel, le logiciel et le média de sauvegarde de remplacement.

3) Phase de mise en place → configurer le matériel du système et installer un OS de base.

4) Phase de restauration → restaurer tout le système à partir du média, y compris les fichiers système et les données utilisateurs.

5) Phase de vérification → vérifier le bon fonctionnement de tout le système et l'intégrité des données utilisateurs.

Indépendamment du SLA, on doit connaître la durée de chaque phase.

De plus, chaque phase peut introduire des retards inattendus !

Par exemple, une phase de diagnostic peut accaparer beaucoup de temps et il est raisonnable de passer à la phase d'approvisionnement si l'anomalie n'est pas trouvée en 15 minutes.

La phase d'approvisionnement peut elle aussi prendre du temps si le média de sauvegarde est hors site et s'il faut attendre sa livraison.

Si le camion chargé de livrer une bande de sauvegarde hors site a un accident en route...

C'est pourquoi il faut toujours un plan d'action et surtout, faire des "répétitions générales" pour vérifier qu'aucun grain de sable ne viendra enrailler le mécanisme de restauration du service