



INSTITUT
INFORMATIQUE
APPLIQUÉE

Sauvegardes et Scripts

Formation présentée par :
Marc-Henri Pamiseux



Être Libre, c'est choisir !

Libricks

Reprise d'activité suite à un incident

- **PRA informatique**

Un Plan de Reprise d'Activité a pour objectif de décrire et de planifier le processus de remise en service d'une infrastructure informatique. Il doit permettre la reprise opérationnelle des services en cas de sinistre. Il est construit sur la base de deux notions importantes :

- RTO : La durée maximale d'interruption admissible (Return Time on Objective)
- RPO : La perte de données Maximale Admissible (Recovery Point Objective)

Pour permettre de qualifier ces deux notions, il faut définir des objectifs de sécurité des données en fonction des besoins élémentaires de l'entité définissant son PRA (Analyse des risques).

RTO (Recovery Time on Objective)

- **Temps maximal acceptable durant lequel une ressource informatique peut rester indisponible suite à un sinistre**
 - Délai pour la détection de l'incident
 - Latence de la décision d'exécuter la procédure de reprise
 - Délai de mise en œuvre du plan de reprise d'activité
 - Configuration du système et du réseau
 - Restauration des données
 - Contrôle de la qualité et du niveau d'efficacité du service (mode dégradé ou pleinement opérationnel)
- **Ce délai est caractérisé par le coût total engendré par une indisponibilité du service.**

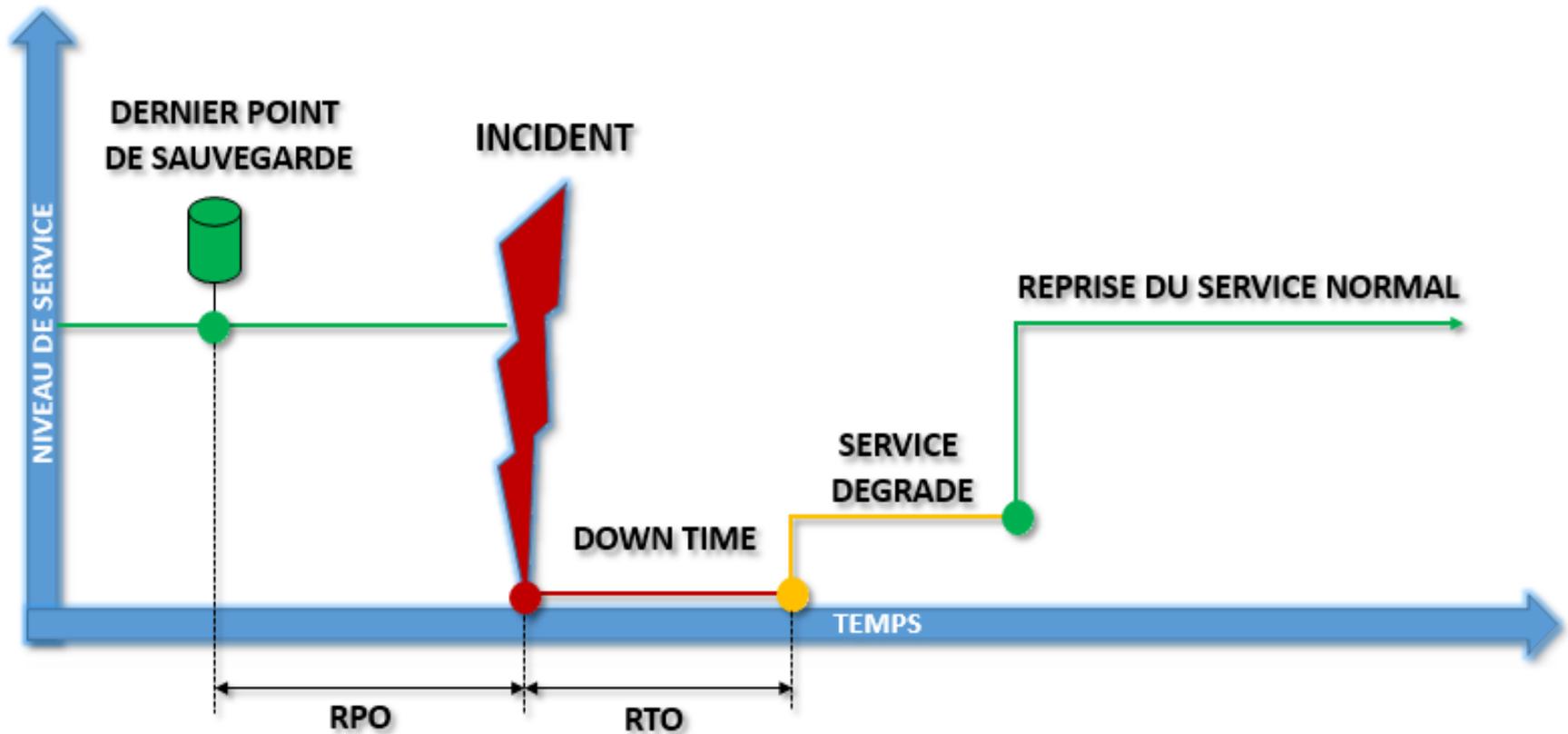
RPO (Recovery Point Objective)

- **Cela définit la quantité maximale de données qui peut être perdue suite à un sinistre informatique.**

En pratique cette quantité est exprimée en minutes ou en heure. Elle correspond au temps écoulé entre la dernière sauvegarde effective et l'incident.

- Interface entre l'utilisateur et la gestion des systèmes informatiques
- Interface avec les fournisseurs et prestataires
- Prise en charge des demandes et des incidents
- Amélioration de la qualité de service
- Connaissance des besoins et des attentes

Schéma d'un incident informatique



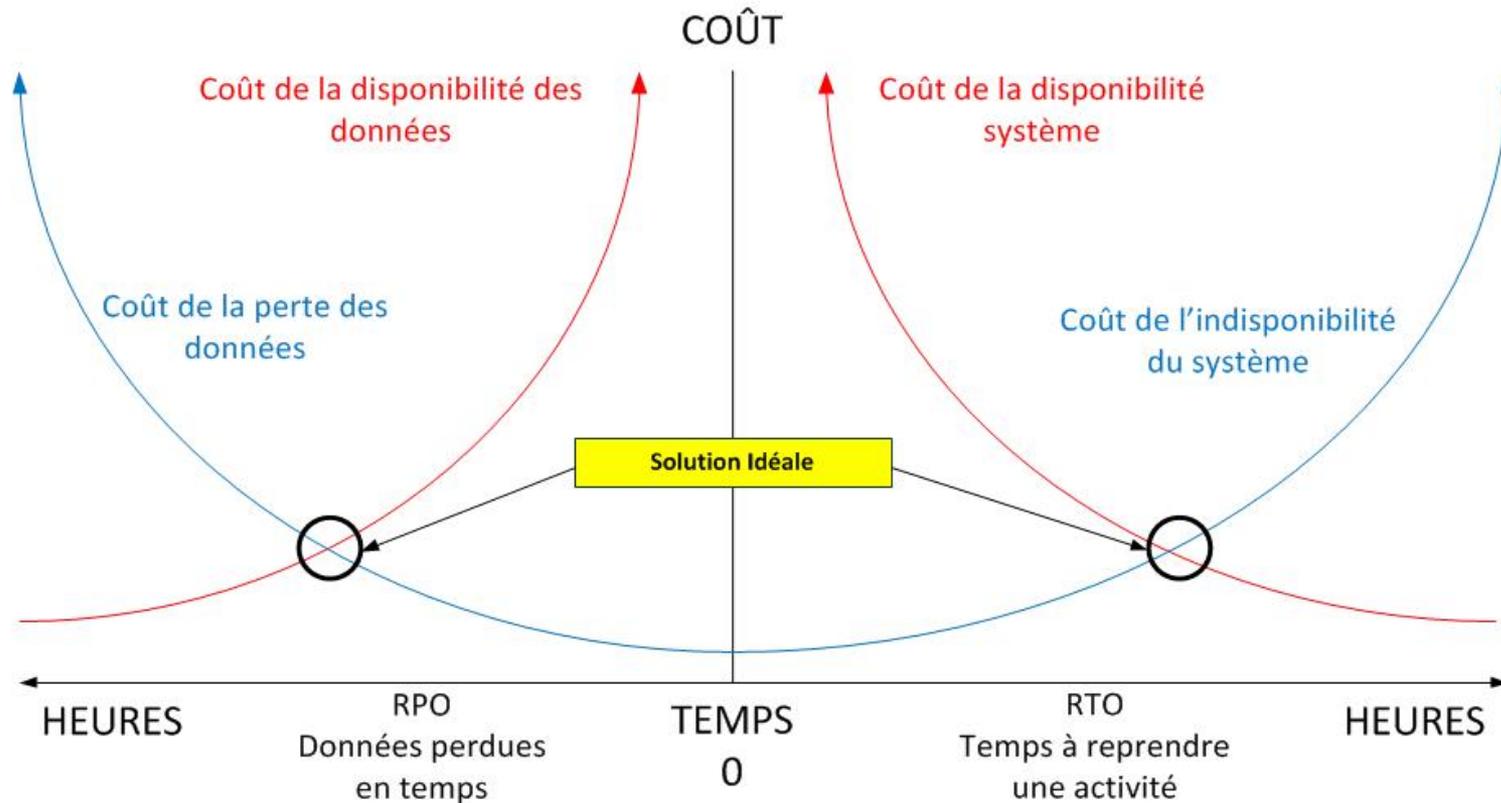
Élaborer un PRA

- **Points de base pour concevoir un PRA**

Processus	Description
Inventaire des actifs informatiques	Recenser tous les équipements actifs, le matériel partagé (serveur, stockage, imprimante...) comme individuel (station de travail, smartphone...), les connectiques (switch, routeurs, câbles...) et qui sont en principe répertoriés dans la CMDB (voir gestion de parc)
Cartographie des applications et des données	Avoir la liste des applications informatiques utilisées par les différents services de l'entreprise et pouvoir localiser les médias d'installation, les licences et surtout les données qu'elles traitent.
Classer les actifs	→ Application vitale : une perte de ce service condamne l'entreprise → Application critique : chômage technique, perte d'exploitation mineure. → Application standard : Peu d'impact majeur sur l'activité de l'entreprise
Prioriser les risques	Définir les règles de priorité de redémarrage des services (suppose une analyse fine des risques).
Conception des seuils RPO/RTO	Confronter les capacités informatiques à faire face à la résolution d'un sinistre par rapport aux besoins et aux risques de l'entreprise.
Analyse des solutions techniques et financières	Assurer un niveau de disponibilité suffisant à un coût raisonnable. Plus les seuils RTO/RPO sont faibles et plus les coûts sont élevés.
Valider/Implémenter/Documenter	Faire valider le PRA par la direction, définir le budget de réalisation et documenter l'ensemble des processus constituant une remise en service.

Coût de la disponibilité et de l'indisponibilité

- Le schéma suivant illustre le lien entre la disponibilité des services, des données et le coût total de production



La sauvegarde des données

- **Enjeu**

Le défi de la sauvegarde consiste à copier de manière sécurisée et horodatée l'ensemble des informations stockées sur un ou plusieurs équipements informatiques en optimisant au maximum l'utilisation de l'espace de stockage.

- **Finalité**

Une copie de sûreté est nécessaire pour restaurer un système informatique dans son état de fonctionnement optimal suite à :

- la perte de fonctionnement du support de stockage
- un incident, une suppression accidentelle
- une modification non souhaitable.

Sauvegarder

- **La sauvegarde**

Elle consiste à copier les informations d'un système sur un support indépendant du système initial.

Cette démarche est formalisée pour l'informatique dans le cadre d'un Plan de Secours (PS) ou plus globalement dans le cadre d'un Plan de Reprise d'Activité (PRA).

Elle peut s'effectuer de manière :

- Manuelle : copie de dossiers
- Semi-automatisée : script ou programme pouvant nécessiter une intervention humaine
- Automatisée : robot de sauvegarde.

Choisir une technique

- **Il est nécessaire de prendre en compte**
 - La capacité du support de stockage (volume du backup).
 - La vitesse de copie vers les supports (durée du backup).
 - La fiabilité des supports en cas de stockage (sensibilité à la température, à l'humidité, à la lumière, à la poussière...)
 - Le volume physique des supports et le coût de l'ensemble.
- **La granularité de la restauration**
 - Pouvoir restaurer un fichier unitairement ou bien toute une archive ?
 - Pouvoir restaurer un enregistrement d'une table et ses dépendances ou toute une base de données ?

Méthodes de sauvegardes

- **Sauvegarde locale (DAS) : disque externe, DAT, DLT, LTO...**
 - Attention à la rotation des médias selon la stratégie utilisée
- **Sauvegarde sur un serveur ou bien un espace de stockage en réseau (NAS ou SAN)**
 - Attention à bien séparer les lieux de stockage pour se prémunir des incendies, des inondations et des vols.
- **Sauvegarde dans le nuage (Cloud)**
 - Attention à bien chiffrer les données externalisées, tant sur la phase de transport que sur la phase de stockage.
 - Attention à contrôler la disponibilité de vos données (quotas, faillite ou différend commercial avec l'hébergeur, malveillance d'un tiers...)

1 - Stratégies de sauvegardes

- **Sauvegarde totale (Full backup)**

Il s'agit de réaliser une copie complète de toutes les données relatives à un service, un équipement ou un système d'information sur un support externalisé à ce système.

Les +

- Image fidèle du SI
- Fonction d'archivage permettant de retrouver une modification à une date précise.

Les -

- Sauvegarde parfois très lente et pouvant ralentir les autres services
- Nécessite un stockage important

2 - Stratégies de sauvegardes

- **Sauvegarde incrémentielle (incremental backup)**

Il s'agit de sélectionner les éléments qui ont été modifiés depuis la sauvegarde précédente qui peut être soit une complète soit une sauvegarde incrémentielle. Il existe aussi un reverse incremental.

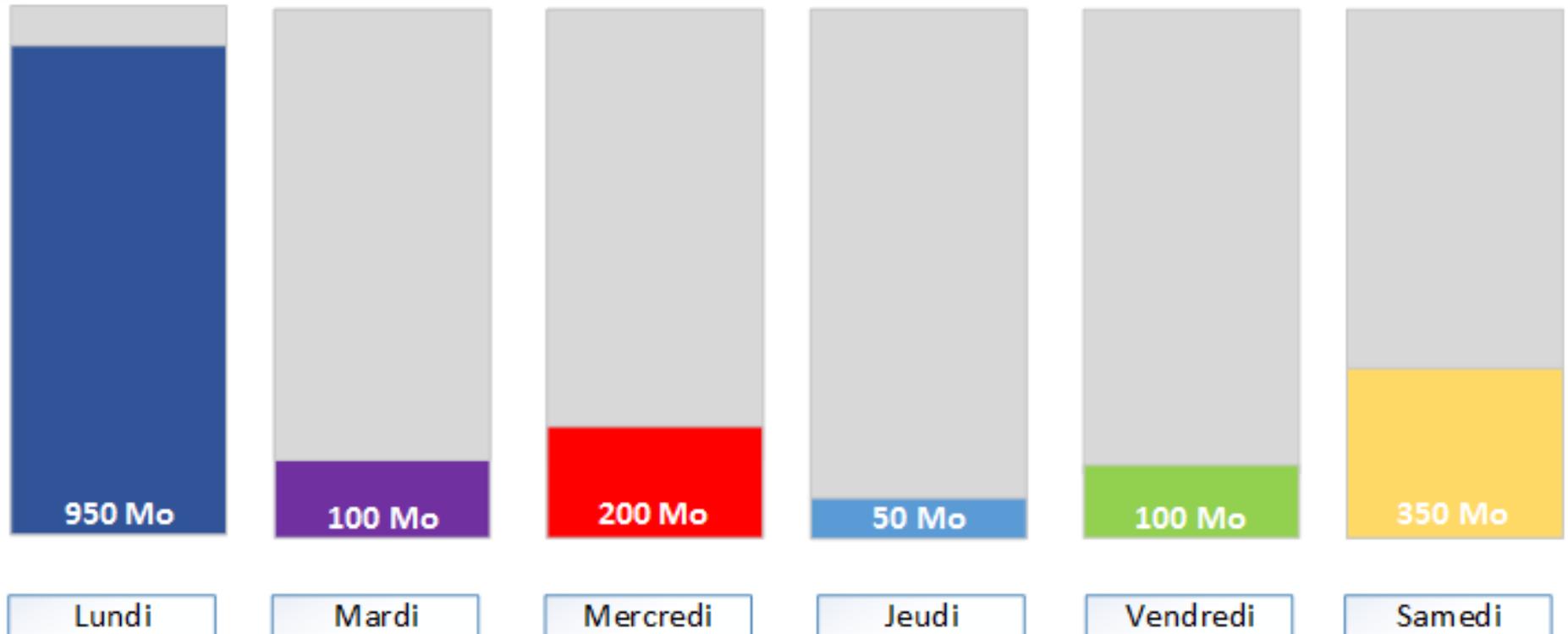
Les +

- Plus rapide car elle se concentre sur un plus petit nombre de données
- L'espace de stockage nécessaire est moindre.

Les -

- En cas de restauration, il faut utiliser l'ensemble des jeux de sauvegardes incrémentielles pour compléter la sauvegarde totale en vue de restituer les données.

2 – Sauvegarde incrémentielle



- **Le stockage nécessaire pour une semaine est de 1750 Mo**

$$\text{Total} = 950 + 100 + 200 + 50 + 100 + 350$$

3 - Stratégies de sauvegardes

- **Sauvegarde différentielle (differential backup)**

On retrouve les principes de la sauvegarde incrémentielle mais on cumule les différences pour chaque sauvegarde. Les images intermédiaires entre la complète et la dernière sauvegarde ne comptent plus.

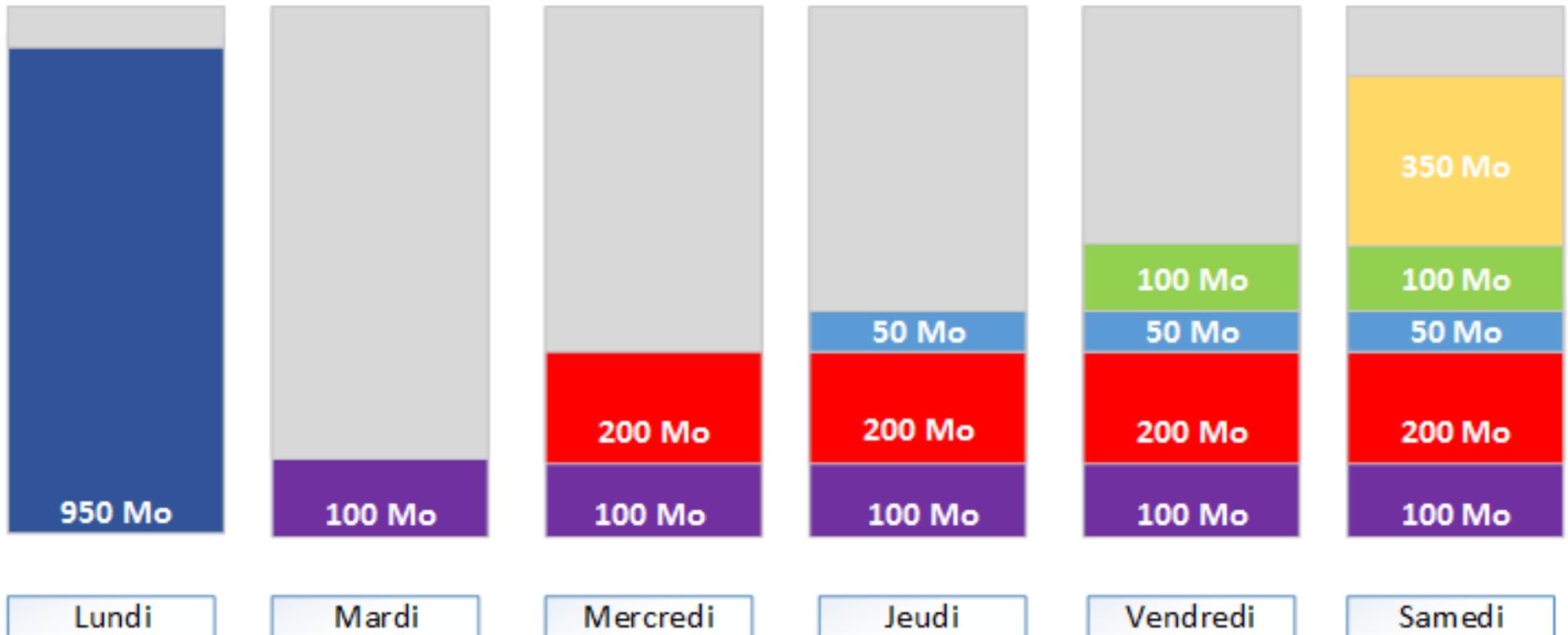
Les +

- Plus rapide et fiable qu'une incrémentielle en cas de restauration.

Les -

- Plus longue à réaliser qu'une sauvegarde incrémentielle
- Nécessite un espace de stockage plus important

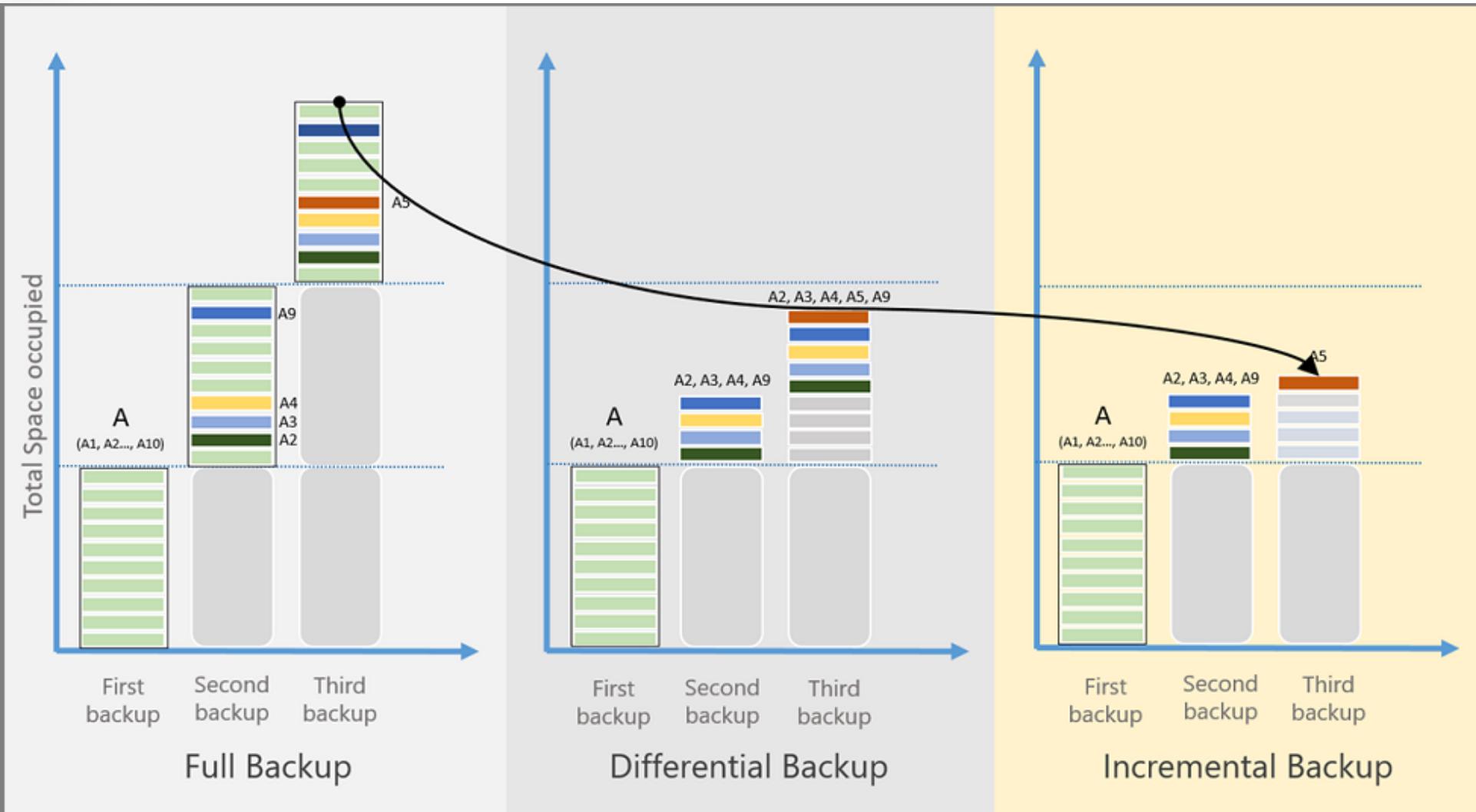
3 – Sauvegarde différentielle



- **Le stockage nécessaire pour une semaine est de 2950 Mo**

$$\text{Total} = 950 + 100 + 300 + 350 + 450 + 800$$

Vue d'ensemble des stratégies



Rétention de sauvegardes

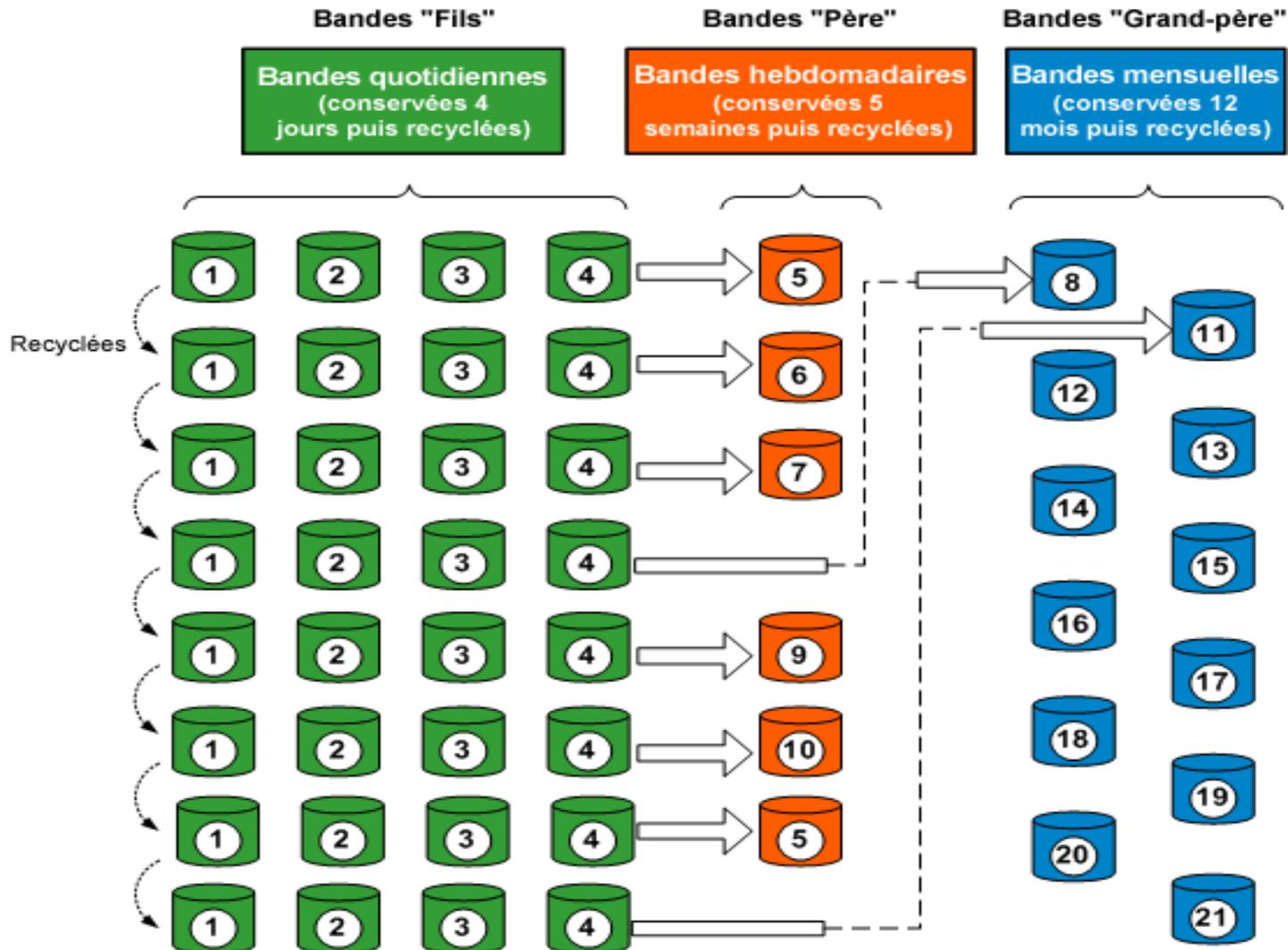
- **La rétention est une règle de conservation**

Une solution de sauvegarde conserve plusieurs versions des données constituant ainsi un historique des modifications. Cet historique se nomme la rétention.

- Rétention par nombre de versions :
 - Chaque sauvegarde est qualifié comme nouvelle version
 - Si le nombre de versions dépasse le nombre de versions autorisées, la plus ancienne est supprimée.
- Rétention temporelle :
 - Ne se base plus sur un nombre de version fini, mais sur des étapes temporelles : Sauvegarde hebdomadaire, mensuelle, annuelle sur 12 mois et 2 ans...

Rotation GFS : Grandfather/Father/Son

Rotation GFS standard sur 5 jours avec 21 bandes



Exemple de rétention de sauvegardes

- **Sauvegarde quotidienne des données**

Sauvegarde complète le lundi puis des sauvegardes différentielles. Une sauvegarde complète le dimanche (J-7).

- **Quatre sauvegardes hebdomadaires par mois**

- On garde J-7, J-14, J-21 et J-28

- **Une sauvegarde mensuelle par an**

- Au delà du 1^{er} mois, il est conservé une sauvegarde par mois pendant un an. On garde donc M-1, M-2, M-3, M-4, M-5, M-6, M-7, M-8, M-9, M-10, M-11, M-12.

- **Un volume de 22 sauvegardes par an :**

- J-1, J-2, J-3, J-4, J-5, J-6, J-7, J-14, J-21, J-28, M-1, M-2, M-3, M-4, M-5, M-6, M-7, M-8, M-9, M-10, M-11, M-12.

Dé-duplication de données

- **Définition**

La dé-duplication consiste à factoriser des séquences de données identiques afin d'économiser l'espace de stockage de ces données.

- Au niveau des fichiers
 - Revient à effectuer un checksum sur chaque fichier à sauvegarder, de le renommer avec cet index puis de stocker ce fichier dans un pool de stockage commun en conservant une table de correspondance entre le fichier unique et son chemin complet initial.
- Au niveau des blocs :
 - Même principe que précédemment mais en découpant les fichiers en bloc de taille fixe et en indiquant ces blocs de manière unique et non plus directement les fichiers.

Règle 3-2-1

- **Comment bien sauvegarder ses données ?**

La règle 3-2-1 permet de diminuer la pertes de données liées à une défaillance des sauvegardes.

- **3 copies de vos données**
 - Comprendre qu'il faut avoir deux copies de sauvegardes en plus de vos données de productions.
- **2 supports différents**
 - Inutile de faire des sauvegardes distinctes si elles se situent sur le même espace de stockage ou sur des équipements situés dans la même pièce.
- **1 copie hors site**
 - Quel que soit le moyen de transfert, il est indispensable d'avoir une copie de sauvegarde délocalisée du site de production.

Sauvegarder n'est pas Archiver

- **La sauvegarde est une protection en cas de désastre**

La sauvegarde est une copie conforme des données de production en vue d'être restaurée en cas d'accident, de pertes ou de vol.

- **L'archivage est un classement des données peu utilisées déplacées en dehors de la production**

L'archive est une démarche de conservation des données qui s'inscrit dans la durée. Ces données restent nécessaires pour une consultation peu fréquente mais ne sont plus modifiées.

- **Les coûts d'archivage doivent être différents des coûts de sauvegarde et plus encore des coûts de stockage.**
 - Veille techno : comparatif des solutions de stockage en ligne selon la fréquence et le type d'accès.
- **Attention à la protection des accès à ces données.**

Liens

- **Forge de Mickaël Bonnard (tuxmika)**

<https://gitlab.com/mickaelbonnard/scripts-bash>

-