

> Ne vous laissez pas duper par la déduplication - Déduplication moderne des données avec Arcserve UDP

par Christophe Bertrand, Vice-Président Marketing Produit

Trop de données, pas assez de temps, pas assez d'espace de stockage et pas assez de budget... Cela vous rappelle quelque chose ? Depuis les premiers mainframes, des efforts ont été faits pour réduire les besoins en matière de capacité de stockage et optimiser les processus de protection des données.

Dans le monde des systèmes ouverts, les problèmes mentionnés ci-dessus sont les mêmes que ceux de l'époque où la première technologie de déduplication des données est devenue un « courant dominant ». Les sauvegardes échouent, prennent trop de place et sont beaucoup trop coûteuses.

Aujourd'hui, les volumes de données ne cessent de croître et les entreprises de toutes tailles peinent à gérer ce qui est devenu un problème très coûteux. Le stockage bon marché aide, mais ce n'est pas une solution efficace sur le plan opérationnel pour de nombreuses charges applicatives. Les données doivent être réduites à des niveaux plus gérables. L'excès de données pose de réels problèmes aux entreprises :

- Les entreprises sur provisionnent leur infrastructure de sauvegarde en prévision d'une croissance future rapide, ce qui est coûteux.
- Les anciens systèmes sont inadaptés et les sauvegardes sont trop longues ou incomplètes.
- Les entreprises n'atteignent pas leurs objectifs de point et de temps de restauration.
- Les sauvegardes surchargent l'infrastructure et la bande passante réseau.
- Les entreprises ne peuvent pas adopter de nouvelles technologies, telles que la sauvegarde Cloud, car la quantité de données à transférer sur les WAN est excessive.

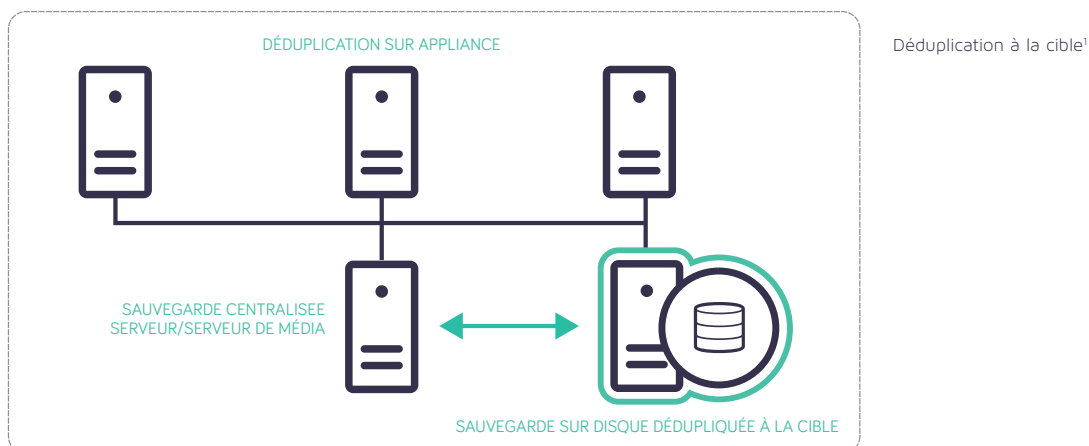
Aujourd'hui, de nouvelles avancées technologiques sont nécessaires pour lutter contre la croissance imparable et exponentielle des machines virtuelles et des données. Dans ce livre blanc, nous allons examiner l'approche Arcserve de la déduplication des données.

Déduplication à la cible... trop has been ?

Au cours des dernières années, les entreprises ont constaté qu'elles ne pouvaient pas respecter leurs fenêtres de sauvegarde avec une architecture de sauvegarde traditionnelle. Outre les problèmes de performance des sauvegardes, elles ont été confrontées au problème du coût du stockage de grandes quantités de données de sauvegarde. Les frais de stockage ont explosé lorsque les schémas de sauvegarde ont implosé. Les appliances de déduplication sont devenues la solution privilégiée pour résoudre ce problème. Le processus consistait à prendre les données de sauvegarde, à les optimiser par le biais de la déduplication et à les stocker sur un disque. Il s'agissait en fait de « compresser » vos volumes de sauvegarde pour réduire vos coûts. La déduplication à la cible a très bien fonctionné. Elle est d'ailleurs encore utilisée aujourd'hui dans de nombreux environnements.

La déduplication à la cible est intéressante, car elle n'oblige pas les utilisateurs à revoir entièrement leurs configurations logicielles ou règles de sauvegarde. Il leur suffit de modifier la destination des flux de sauvegarde.

La déduplication à la cible s'exécute à la volée ou en post-traitement (toutes les données sont écrites sur le disque, puis optimisées a posteriori pour accélérer les sauvegardes).



Dans l'une de ses sessions pédagogiques de 2008, la SNIA présente des scénarios de déduplication à la cible. Dans le cas de la déduplication à la cible, on voit que le logiciel de sauvegarde fonctionne en tant que Data Mover et envoie tous les flux de données de sauvegarde non dédupliqués au disque cible ou à l'appliance VTL.

Que le traitement soit effectué en mémoire à la volée ou après la sauvegarde, pourquoi payer pour une couche technologique supplémentaire ? C'est une approche et une technologie du passé. Aujourd'hui, il existe une meilleure solution.

Déduplication distribuée... si coûteuse

Certains fournisseurs augmentent les coûts en facturant chaque agent de déduplication distribuée sur chaque système à protéger. Malgré d'éventuels avantages ou améliorations du point de vue de la reprise après sinistre, la multiplication de systèmes coûteux et de licences logicielles, ainsi que les besoins associés en matière de bande passante, ne font qu'augmenter la note. En outre, malgré la disponibilité éventuelle d'outils, cette architecture peut s'avérer coûteuse sur le plan de la gestion.

Déduplication à la source... Nouvelle génération

Pourquoi dédupliquer les données a posteriori si vous pouvez sauvegarder uniquement les données nouvelles et uniques requises à la source ? Tant que vous n'impactez pas le client, vous pouvez économiser la bande passante que vous utilisez pour envoyer les flux de sauvegarde à la cible.

Pourquoi ne pas partager toutes ces informations dédupliquées avec tous vos clients ? C'est ce que l'on appelle la « déduplication globale à la source » et c'est la direction que doit prendre le marché. La déduplication globale à la source est ce que de nombreux utilisateurs considèrent comme la technologie de sauvegarde critique pour l'avenir. Le défi de la déduplication globale à la source est de s'assurer que le système source (client) n'est pas saturé par le logiciel de déduplication.

¹Source : http://www.snia.org/sites/default/education/tutorials/2008/spring/data-management/Hamilton-D_Deduplication_Methods_Data_Efficiency.pdf



Arcserve Unified Data Protection

Arcserve Unified Data Protection (UDP) permet aux administrateurs de la sauvegarde de réduire les besoins en matière de stockage et de bande passante, tout en accélérant la protection et la restauration des données sur plusieurs sites. En outre, cette solution permet la réhydratation sur place des données pour une restauration granulaire rapide, y compris à partir d'une bande.

Arcserve UDP effectue une déduplication globale à la source, sur le ou les nœuds protégés. Il transmet seulement les données du nœud source qui sont uniques au serveur UDP RPS (Recovery Point Server) qui fonctionne en tant que datastore central. Après une première sauvegarde instantanée complète, tous les autres sauvegardes capturent uniquement les modifications incrémentielles des données. Ce processus côté client étant très efficace, il en résulte les avantages suivants :

- Réduction de la quantité de données sauvegardées
- Réduction de la quantité de données transmises sur le réseau
- Réduction de la charge du serveur de production et possibilité d'augmenter la fréquence des sauvegardes pour améliorer le RPO et la protection du système, des applications et des données.

Avec la déduplication globale à la source d'Arcserve UDP, vous ne sauvegardez jamais les données deux fois, y compris les systèmes d'exploitation, hyperviseurs, etc. La déduplication globale à la source d'Arcserve UDP réduit considérablement la quantité de données transférées pendant les cycles de sauvegarde. Cela s'applique aux logiciels UDP et aux appliances UDP qui exécutent les mêmes logiciels.

Pour résumer, chaque ordinateur, machine virtuelle ou serveur sauvegardé communique avec le serveur Arcserve UDP Recovery Point Server (RPS) qui gère, dans sa base de données globale, un index des fichiers de toutes les machines, où qu'elles se trouvent. Le serveur RPS se charge de déterminer ce qui doit être sauvegardé et extrait les nouvelles données tout en éliminant les copies redondantes. Couplé à une technologie de compression, RPS peut réduire de plus de 92 % les données stockées sur le disque.

L'index de la base de données de déduplication de RPS est stocké sur un disque SSD hautes performances, afin d'optimiser les performances et l'efficacité tout en réduisant les coûts par rapport à une approche basée uniquement sur la mémoire.

Destinations: Recovery Point Server							
Actions - Add a Recovery Point Server							
	Name	Plan Count	Data Protected	Deduplication	Compression	Overall Data Reduction	Space Occupied
+	...						
✓	...	6	15.57 TB	61%	34%	74%	4.02 TB
✓	...	5	14.41 TB	54%	39%	72%	4.06 TB
✓	...	1	6.24 TB	20%	40%	56%	2.72 TB
✓	...	6	11.78 TB	89%	37%	92%	936.78 GB
✓	...	1	94.36 GB	66%	37%	79%	20.41 GB
!	...	1	N/A	N/A	N/A	N/A	N/A
+	...						
✓	...	4	12.85 TB	64%	33%	78%	2.94 TB
✓	...	3	5.59 TB	70%	33%	84%	915.21 GB
+	...						
✓	...	1	6.25 TB	20%	40%	56%	2.72 TB
✓	...	4	7.61 TB	66%	39%	81%	1.47 TB
✓	...	5	9.69 TB	89%	29%	92%	757.4 GB

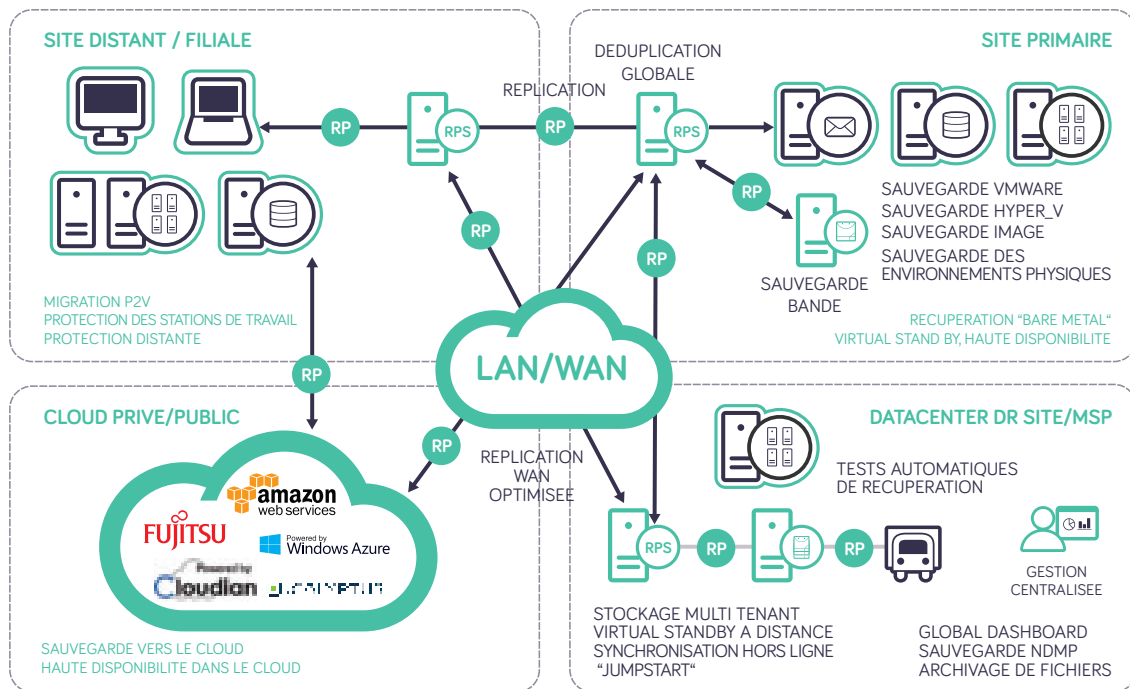
Cette capture d'écran réelle d'un client montre l'efficacité de la déduplication combinée avec les fonctions de compression supplémentaires que nous appliquons au datastore, ce qui se traduit par des niveaux de réduction des données généraux très élevés.



Déduplication réellement globale... Impressionnante

La déduplication globale couvrant tous les clients au sein de l'infrastructure est essentielle pour limiter le stockage et le transfert inutiles des données de sauvegarde redondantes. Les données sont dédupliquées sur tous les nœuds, dans toutes les procédures et sur tous les sites. La déduplication globale d'Arcserve UDP dépasse les limites de certains fournisseurs dont la déduplication ne s'applique qu'au cache de réplication du WAN, et non à ce qui est réellement stocké sur le disque, ce qui réduit les avantages potentiels globaux en termes d'économies sur la bande passante et le stockage.

Pour la protection des données sur les sites locaux et distants (Cloud privé), l'index de la base de données de déduplication globale d'Arcserve UDP est répliqué et distribué afin que toutes les données sources et cibles soient dédupliquées sur tous les serveurs RPS. Les données de sauvegarde étant dédupliquées globalement avant leur transfert vers le serveur RPS cible, seules les modifications sont envoyées sur le réseau, ce qui améliore les performances et réduit la consommation de bande passante. L'ensemble de ce processus est sécurisé par le chiffrement au niveau du datastore et par les mots de passe de session.



Architecture de déduplication globale d'Arcserve UDP



Le résultat

La déduplication d'Arcserve UDP n'est pas seulement une meilleure technologie. Elle est également porteuse d'avantages pour les entreprises :

- **Sauvegardes complètes plus rapides** : la réduction des données à transmettre et à stocker accélère les sauvegardes. C'est important lorsque le volume total des données risque d'être long à sauvegarder au point qu'une sauvegarde ne soit pas terminée avant le début de la suivante.
- **Amélioration des performances des clients** : l'agent Arcserve a un impact minimal sur les performances des clients, car la majorité du traitement s'exécute sur le serveur RPS. Pour les environnements virtualisés, la sauvegarde VMware et Hyper-V sans agent réduit le risque de goulots d'étranglement et de problèmes de performance au niveau du matériel. Autrement dit, elle garantit que les sauvegardes ne bloqueront pas les serveurs virtualisés.
- **Infrastructure de sauvegarde simplifiée** : le routage des sauvegardes via le serveur RPS facilite la transmission des données de sauvegarde vers différentes destinations, par exemple vers le Cloud via les services Azure ou AWS (Amazon Web Services) ou vers une bande locale ou un Cloud privé.
- **Amélioration de la résilience et de la disponibilité** : la centralisation du stockage des données dédupliquées sur le serveur RPS facilite la protection de l'infrastructure de sauvegarde. Par exemple, toutes les informations du datastore peuvent être répliquées vers un Cloud ou un autre serveur dans le même datacenter ou hors site.
- **Réduction de la bande passante nécessaire pour les sauvegardes** : le serveur RPS extrait uniquement les données nouvelles ou modifiées des clients, avec de hauts niveaux de granularité (jusqu'aux segments de 4 Ko). Il en résulte des sauvegardes extrêmement efficaces.
- **Respect des engagements contractuels concernant le RPO et le RTO** : la déduplication globale à la source d'Arcserve facilite la réalisation des objectifs de point de restauration et de temps de restauration. La réduction des données à transférer permet d'augmenter la fréquence des sauvegardes et d'accélérer les restaurations.

Jugez-en par vous-même ! Testez Arcserve UDP gratuitement pendant 30 jours : rendez-vous sur arcserve.com/free-backup-software-trials

Consultez notre outil de calcul de la rentabilité de la déduplication UDP : arcserve.com/calculator

arcserve[®]
assured recovery™

Pour en savoir plus sur Arcserve UDP, rendez-vous sur arcserve.com