

Livre blanc sur la déduplication des données

Basé sur Active Backup for Business 2.3.0



Table des matières

Introduction	3
Technologie de réduction des données	4
Processus de déduplication	
Changed Block Tracking (CBT)	
Taux de déduplication	
Conclusion	9

Où trouver les informations dont vous avez besoin

Synology a publié un large éventail de documents de support.

Notre [Centre de connaissances](#) contient des articles d'aide et de FAQ utiles, ainsi que des tutoriels vidéo qui décomposent les processus en étapes pratiques. Vous y trouverez également des guides de l'utilisateur, des guides de solutions, des brochures et des livres blancs. Les utilisateurs expérimentés et les administrateurs trouveront des réponses et des conseils techniques dans les guides de l'administrateur et les guides du développeur.

Vous avez un problème et vous ne trouvez pas la solution dans notre documentation officielle ? Effectuez une recherche parmi des centaines de réponses publiées par des utilisateurs et le personnel d'assistance dans [Synology Community](#) ou contactez l'[assistance Synology](#) par le biais du formulaire Web, par e-mail ou par téléphone.



Introduction

La perte de données ou la défaillance d'un système peut être désastreuse, quel que soit le périphérique ou le système de stockage que vous utilisez. Les périphériques les plus importants pour vous, tels que vos ordinateurs ou serveurs professionnels, méritent une stratégie de sauvegarde concrète et efficace qui peut vous permettre d'éviter les pannes système inattendues en cas de catastrophes naturelles ou d'autres problèmes techniques.

Active Backup for Business (ABB) vous permet de gérer toutes vos sauvegardes de périphérique VMware, Hyper-V, Windows, Linux et de serveur de fichiers à partir d'une console centrale unique sur votre Synology NAS. Grâce aux options de restauration instantanée pour les fichiers individuels, les machines virtuelles et les périphériques Windows, vous pouvez assurer le fonctionnement des services critiques, même en cas de sinistre soudain. En outre, Active Backup for Business utilise un certain nombre de techniques les plus récentes, telles que la **déduplication des données** et la **sauvegarde incrémentale**, pour améliorer les performances de sauvegarde et réduire la quantité d'espace physique requise par les données de sauvegarde. Ces approches étant si couramment utilisées par les différents services de sauvegarde, nos utilisateurs se renseignent fréquemment sur les technologies et les mesures utilisées pour obtenir le taux de déduplication d'ABB.

L'objectif de ce document est de présenter certaines des techniques de déduplication de données utilisées et de définir le taux de déduplication mis en œuvre dans Synology Active Backup for Business.

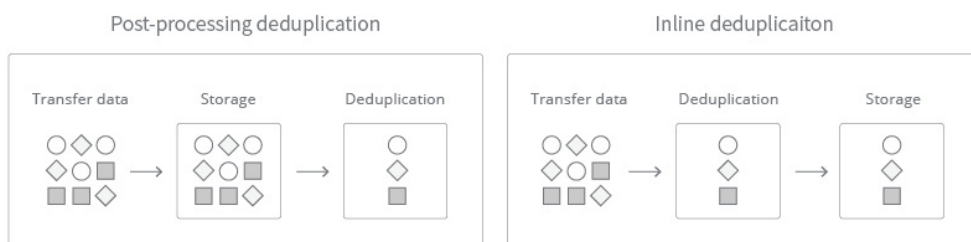
Technologie de réduction des données

La technologie de réduction des données est conçue pour réduire l'espace de stockage utilisé par les données de sauvegarde, en optimisant l'utilisation du stockage et en permettant à l'utilisateur de sauvegarder autant de périphériques que possible. Active Backup for Business a mis en œuvre une technologie de pointe pour réduire les données et améliorer les performances globales de sauvegarde.

Processus de déduplication

La déduplication est un moyen efficace d'optimiser l'espace de stockage en supprimant les données redondantes. Le processus de déduplication peut être réparti en deux catégories : « en ligne » et « post-traitement ». La **déduplication en ligne** supprime toutes les données redondantes **avant** qu'elles ne soient écrites dans le stockage, tandis que la **déduplication post-traitement** supprime les données **après** la fin du processus de sauvegarde.

La déduplication en ligne peut être mise en œuvre à la fois sur l'agent de sauvegarde et sur la destination de sauvegarde, selon la solution de sauvegarde. Cependant, lorsque la déduplication est mise en œuvre via un agent, l'analyse, la déduplication et la transmission des données se produisent toutes en même temps, ce qui se traduit par un processus de sauvegarde plus long et moins performant (destination de sauvegarde).



Synology **Active Backup for Business** intègre une fonction de **déduplication en ligne** dans sa conception. Cependant, les processus de déduplication des données se produisent uniquement sur la destination avant l'écriture sur le stockage. Cela permet d'éviter de consommer trop de temps et de ressources CPU sur la source de sauvegarde, un problème qui est généralement causé par une déduplication côté source. En outre, Active Backup for Business permet aux utilisateurs de sauvegarder divers systèmes et périphériques sur une seule destination pouvant contenir des données redondantes.

Introduction

Afin d'éliminer toutes les copies redondantes de données sur plusieurs cibles, Active Backup for Business offre également une fonction de **déduplication globale**. La déduplication globale permet à Active Backup for Business de dédupliquer les données de nombreuses tâches de sauvegarde au sein d'un seul dossier partagé, optimisant ainsi la consommation réelle du stockage sans compromettre les performances globales de sauvegarde.

En ce qui concerne le mécanisme de déduplication, Active Backup for Business utilise un identifiant unique pour classer chaque fichier et fragment de fichier sauvegardé. Pour ce faire, le moteur de déduplication d'Active Backup for Business divise les données de sauvegarde en fragments de fichiers individuels de **4Ko** et attribue à chacun une empreinte unique. Cette empreinte est calculée par un moteur de déduplication qui utilise l'**algorithme SHA-256** pour identifier chaque fragment de données.

Une fois qu'une empreinte a été créée pour un fragment de fichier, le système crée un index des empreintes pour chaque fragment de fichier stocké. Cet index est utilisé pour faire correspondre les fragments de fichiers entrants avec leur homologue stocké. Une fois identifiées, le moteur de déduplication supprime rapidement les données redondantes, optimisant ainsi l'utilisation du stockage du périphérique.

Changed Block Tracking (CBT)

Le **Suivi des blocs modifiés (CBT)** est un type de technologie de sauvegarde incrémentale couramment utilisée par un certain nombre de solutions de sauvegarde. La **sauvegarde incrémentale** est un type de sauvegarde qui ne sauvegarde que les données qui ont été modifiées ou récemment créées depuis la version de sauvegarde la plus récente. La fonction CBT facilite les sauvegardes incrémentales en économisant du temps et des ressources, en ne sauvegardant que les blocs de données modifiés plutôt qu'en effectuant une sauvegarde complète à chaque fois. Pour cette raison, la taille des données qui seront transférées via l'application de sauvegarde, c'est-à-dire la **taille des données de sauvegarde transférées**, est inférieure à la taille des données d'origine sur la source de sauvegarde.

L'architecture d'Active Backup for Business inclut cette fonction pour les sauvegardes sur PC, serveurs physiques et machines virtuelles. ABB effectue une sauvegarde incrémentale permanente pour une tâche de sauvegarde afin d'optimiser le nombre de versions de sauvegarde disponibles et de réduire la quantité de stockage utilisée pour la conservation des sauvegardes en sauvegardant uniquement les données modifiées après l'exécution d'une sauvegarde complète.



Taux de déduplication

Définition

Le **taux de déduplication** est un indicateur qui permet d'évaluer l'efficacité des performances de déduplication de données d'une solution de sauvegarde. Ce taux est souvent un facteur décisif lorsque des utilisateurs choisissent une solution de sauvegarde, car il leur permet d'analyser les différentes fonctions et performances offertes par divers fournisseurs. Cependant, les taux de déduplication peuvent varier en fonction de la technologie de déduplication utilisée et des caractéristiques générales des données.

Un **taux de déduplication** ou **taux de compression** est défini comme la mesure des données qui doivent être stockées et la quantité réelle d'espace de stockage utilisé pour stocker les données de sauvegarde sur la destination de sauvegarde après la suppression de données redondantes.

Par exemple : vous avez sauvegardé 100 Go de données à partir de votre périphérique et la quantité réelle d'espace utilisé sur le serveur de stockage de sauvegarde était de 20 Go. Si vous divisez la quantité de données de sauvegarde (100 Go) par l'espace utilisé sur le serveur de stockage de sauvegarde (20 Go), vous obtenez un taux de déduplication de 5:1, soit un taux de déduplication de 80 %.

Le taux de déduplication est également largement utilisé dans le domaine du marketing pour promouvoir les points forts d'un produit de sauvegarde. Par exemple, les campagnes promotionnelles peuvent indiquer qu'un produit de sauvegarde offre un taux de déduplication 20 fois supérieur à celui d'autres produits disponibles sur le marché. Cependant, les taux de déduplication varient en fonction de la technologie de déduplication utilisée, des caractéristiques générales des données, des méthodes de sauvegarde utilisées et de la manière dont le taux de déduplication est calculé.

Pour cette raison, le calcul du taux de déduplication n'est pas toujours le moyen le plus précis de comparer différents produits. Par conséquent, il est important de garder à l'esprit la manière dont le taux de déduplication de chaque produit est calculé, ainsi que la quantité de stockage correspondante réellement utilisée sur la destination de sauvegarde.

Calcul

Le taux de déduplication peut être calculé à l'aide des valeurs suivantes :

- la capacité totale des données de sauvegarde avant la suppression des doublons ;
- la taille des données de sauvegarde transférées ;
- la capacité réelle utilisée sur la destination de sauvegarde.

En général, une bonne méthode de calcul du taux consiste à diviser la **taille des données de sauvegarde transférées** par la **capacité réelle utilisée sur la destination de sauvegarde**. Il s'agit de la méthode utilisée par Synology.

Introduction

En revanche, certains fournisseurs de sauvegarde divisent la **capacité totale des données de sauvegarde avant la suppression des doublons** par la **capacité réelle utilisée sur la destination de sauvegarde**. Cela ne permet cependant pas nécessairement d'obtenir un taux de déduplication précis.

Prenons, par exemple, le scénario suivant :

- Solution de sauvegarde : A
- Capacité totale des données de sauvegarde avant la suppression des doublons : 6 To
- Taille des données de sauvegarde transférées : 1,2 To
- Capacité réelle utilisée sur la destination de sauvegarde : 0,83 To
- Taux de déduplication annoncé : 7,2:1

Dans ce scénario, la solution de sauvegarde A est annoncée comme ayant un taux de déduplication de **7,2:1** (6 To/0,83 To). Cependant, ce calcul n'est pas tout à fait exact. Cela est dû au fait que seul **1,2 To de données de sauvegarde** a été transféré, que la quantité de données redondantes a ensuite été réduite via la fonction CBT et que les données ont été compressées à **0,83 To** pour les stocker sur la destination de sauvegarde.

La **solution de sauvegarde A** divise la **capacité totale des données de sauvegarde** par la **capacité réelle utilisée sur la destination de sauvegarde** afin d'obtenir un taux de déduplication plus élevé. Mais en réalité, seule la **taille des données de sauvegarde transférées** indique la quantité réelle de données qui seront dédupliquées.

Si l'on divise la **taille des données de sauvegarde transférées** par la **capacité réelle utilisée sur la destination de sauvegarde**, on obtient un résultat encore plus précis. Par conséquent, les calculs de déduplication peuvent être ajustés comme suit :

Taux de déduplication

Taille des données de sauvegarde transférées	÷	Capacité réelle utilisée sur la destination de sauvegarde	=	Taux de déduplication des données
1,2 To		0,83 To		1,44:1

Taux de déduplication

(Taille des données de sauvegarde transférées	-	Capacité réelle utilisée sur la destination de sauvegarde)	÷	Taille des données de sauvegarde transférées	x 100 %	=	Taux de déduplication
	1,2 To		0,83 To			1,2 To			30,8 %

Méthode de sauvegarde

Un autre élément est susceptible d'influencer le taux de déduplication : la **méthode de sauvegarde**, notamment si les sauvegardes sont toutes des **sauvegardes complètes** ou si une **sauvegarde incrémentale** est utilisée. Les systèmes de sauvegarde qui effectuent uniquement des sauvegardes complètes peuvent sembler offrir un meilleur taux de déduplication que ceux qui effectuent des sauvegardes incrémentales. Cela vient du fait que les sauvegardes incrémentales transfèrent uniquement les blocs de données modifiés et suppriment toutes les données redondantes avant de les transférer, ce qui réduit la capacité réelle utilisée sur la destination de sauvegarde. Les sauvegardes complètes, quant à elles, incluent une grande quantité de données redondantes par rapport à la version de sauvegarde précédente ou à la version d'origine.

En résumé, bien que les solutions de sauvegarde incrémentale n'affichent pas un meilleur taux de déduplication que les solutions de sauvegarde complète, elles consomment tout de même moins d'espace de stockage réel sur la destination de sauvegarde.

Taux de déduplication d'Active Backup for Business

Comme mentionné précédemment, si la **capacité totale des données de sauvegarde avant la suppression des doublons** est utilisée pour calculer le taux de déduplication, il est possible d'obtenir de meilleurs taux de déduplication, mais ceux-ci ne reflètent pas nécessairement la quantité réelle de données dédupliquées. Active Backup for Business utilise plutôt la **taille des données de sauvegarde transférées**, car elle représente la quantité réelle de données qui vont être dédupliquées. Comme indiqué dans le tableau ci-dessous, les techniques de déduplication d'Active Backup for Business entraînent une réduction de la **capacité réelle utilisée sur la destination de sauvegarde**, et permettent donc aux utilisateurs d'économiser davantage d'espace de stockage dans l'ensemble.

En conclusion, au lieu de s'appuyer sur les taux de déduplication des produits, les utilisateurs doivent se focaliser davantage sur la quantité réelle de stockage utilisée par rapport à la taille du transfert des sauvegardes pour mieux comprendre comment le produit permet d'économiser de l'espace de stockage.

Le tableau suivant indique le taux de déduplication réel d'**Active Backup for Business** et le compare aux taux de déduplication des autres fournisseurs de sauvegarde.

Fournisseur	Taille des données de sauvegarde transférées	Capacité réelle utilisée sur la destination de sauvegarde	Taux de déduplication	Taux de déduplication
Synology Active Backup for Business	1,2 To	637,2 Go	$1,2 \text{ To} / 0,62 \text{ To} = 1,93:1$	48,18 %
Fournisseur A	1,2 To	819,8 Go	$1,2 \text{ To} / 0,8 \text{ To} = 1,5:1$	33,33 %
Fournisseur B	1,2 To	847,7 Go	$1,2 \text{ To} / 0,82 \text{ To} = 1,45:1$	31,08 %
Contenu de la source de sauvegarde : 9 PC Windows, Windows Server sur 1 machine virtuelle				



Conclusion

La déduplication des données est un processus complexe qui permet aux utilisateurs de réduire la quantité de données redondantes afin de gérer plus efficacement les sauvegardes et d'optimiser l'espace de stockage. Étant donné que la déduplication est une technique couramment utilisée par un certain nombre de solutions de sauvegarde, de nombreux utilisateurs souhaitent connaître les taux de déduplication afin de comparer différents produits. Cependant, un certain nombre de méthodes de mesure sont disponibles et différents facteurs sont susceptibles d'avoir un impact sur les techniques qui permettent de mesurer la déduplication. Les taux obtenus pour chaque produit ne sont donc pas toujours fiables. Par conséquent, lorsque vous étudiez le taux de déduplication de différents produits, il est plus important de vous concentrer sur la quantité réelle d'espace de stockage physique utilisée pour stocker les données de sauvegarde après la déduplication.

Synology Active Backup for Business offre non seulement une solution de sauvegarde complète qui peut être utilisée sur diverses plateformes, mais dispose également d'une architecture qui permet de dédupliquer les données rapidement, efficacement, et de manière fiable ce qui vous permet de gagner du temps et de tirer le meilleur parti de votre espace de stockage.



**SYNOLOGY
INC.**

9F, N° 1, Yuan Dong Rd.
Banqiao, New Taipei 220545
Taiwan
Tél. : +886 2 2955 1814

**SYNOLOGY
AMERICA CORP.**

3535 Factoria Blvd SE, Suite #200,
Bellevue, WA 98006
États-Unis
Tél. : +1 425 818 1587

**SYNOLOGY
UK LTD.**

Unit 5 Danbury Court, Linford Wood,
Milton Keynes, Bucks, MK14 6PL,
Royaume-Uni
Tél. : +44 (0)1908048029

**SYNOLOGY
FRANCE SARL**

102 Terrasse Boieldieu (TOUR W)
92800 Puteaux
France
Tél. : +33 147 176288

**SYNOLOGY
GMBH**

Grafenberger Allee 295
40237 Düsseldorf
Allemagne
Tél. : +49 211 9666 9666

**SYNOLOGY
SHANGHAI**

200070, Room 201,
N° 511 Tianmu W. Rd.,
Jingan Dist., Shanghai,
Chine

**SYNOLOGY
JAPAN CO., LTD.**

4F, N° 3-1-2, Higashikanda
Chiyoda-ku Tokyo, 101-0031
Japan

Synology®



synology.com

Synology se réserve le droit de modifier les spécifications et les descriptions de ses produits à tout moment et sans préavis. Copyright © 2022, Synology Inc. Tous droits réservés. ® Synology et les autres noms des produits Synology sont des marques commerciales ou des marques commerciales déposées de Synology Inc. Les autres noms de produits et de sociétés mentionnés dans ce document sont des marques commerciales de leurs propriétaires respectifs.