

LIVRE BLANC

Sauvegarde et restauration : amélioration de l'efficacité et réduction des coûts informatiques grâce à la déduplication des données

Parrainé par : EMC Corporation

Laura DuBois

Robert Amatruda

Février 2010

RÉSUMÉ ANALYTIQUE

La déduplication des données représente une amélioration considérable en matière de maîtrise des coûts informatiques. En effet, elle permet de réduire les besoins en espace de stockage, les fenêtres de sauvegarde et la consommation de bande passante réseau sur les sites distribués comme dans les datacenters. Dans les environnements réels, la déduplication améliore l'efficacité des sauvegardes et des restaurations et réduit les coûts informatiques. Ce livre blanc décrit les diverses méthodes de déduplication des données de sauvegarde et indique les facteurs à prendre en compte lors du choix d'une solution. Il présente également l'offre de solutions de déduplication EMC pour la sauvegarde et la restauration, ainsi que des exemples d'utilisation spécifiques pour une efficacité et une rentabilité optimale des sauvegardes.

Le choix de la déduplication

Face à des besoins en stockage qui doublent pratiquement chaque année, les entreprises de taille intermédiaire et les grands comptes sont de plus en plus attirés par les technologies de déduplication des données. Cette croissance des volumes de données est due aux nouvelles applications, à la généralisation de la virtualisation, à la création de banques de documents électroniques, au partage des documents, à l'utilisation des technologies Web 2.0, et à la rétention ou conservation des enregistrements numériques. Parallèlement aux restrictions des budgets informatiques, la nécessité de freiner les besoins en stockage de données est d'autant plus forte que les entreprises cherchent à réduire leurs coûts d'investissement et d'exploitation. Du point de vue de l'infrastructure physique, de nombreux responsables de datacenter doivent également faire face à des limites en termes d'alimentation, de ventilation et d'espace. La déduplication est une technologie qui non seulement contribue à améliorer l'efficacité du stockage en réduisant les coûts, mais allège également la charge sur les datacenters physiquement limités.

La déduplication apporte en outre une réponse aux problèmes d'efficacité en matière de gestion, de sauvegarde et de réseau. Avec l'explosion actuelle des données, on constate un rapport de plus en plus disproportionné entre l'effectif informatique et le volume d'espace de stockage à gérer. La déduplication réduit le volume de données, permettant ainsi d'équilibrer ce rapport. De même, alors que le fossé entre la puissance de traitement des serveurs et les disques ne cesse de se creuser, les entreprises cherchent à améliorer les performances dans l'ensemble de leur environnement, que ce soit au niveau du WAN, des sous-systèmes de stockage de disque ou des fenêtres de sauvegarde. La technologie de déduplication des données optimise l'infrastructure physique et

virtuelle en transférant moins de données via les liaisons réseau locales ou distantes. Elle réduit également les temps de réponse et les fenêtres de sauvegarde. La déduplication utilise par ailleurs des supports à accès aléatoire (disque), améliorant ainsi les délais de restauration, la sécurité des données et la fiabilité.

De nouveaux défis ont émergé récemment avec le développement de la virtualisation. À mesure que les entreprises déploient la technologie des machines virtuelles à des fins de consolidation des serveurs et de reprise après sinistre, les machines virtuelles traitent des données qui doivent être protégées, même si elles peuvent être très redondantes. Pour parer les différents scénarios de défaillance ou pour restaurer une image, il est généralement nécessaire d'intégrer un serveur physique et des fichiers distincts dans une solution et un processus de sauvegarde uniques. La déduplication permet de réduire de façon significative la capacité de stockage requise pour les sauvegardes, car elle peut éliminer la redondance généralement présente dans les fichiers VMDK. Cependant, les méthodes standard, telles que le déploiement d'un agent de sauvegarde traditionnel dans une machine virtuelle invitée ou le recours à une sauvegarde par proxy VCB pour créer une sauvegarde au niveau des images, ne réduisent pas le volume de données des machines virtuelles à sauvegarder, ni la bande passante de réseau local requise pour déplacer les données. Alliée à un logiciel de sauvegarde, la déduplication constitue une solution complète, efficace et économique de protection des environnements de machines virtuelles.

Avantages de la déduplication

Les entreprises mettent en oeuvre la déduplication à différents niveaux de la pile d'infrastructure afin de relever les défis auxquelles elles sont confrontées. La déduplication présente notamment les avantages suivants :

- ☒ **Réduction des coûts** : la déduplication optimise les ressources et réduit les coûts car elle permet notamment de diminuer les besoins en alimentation, ventilation et espace dans les datacenters, ainsi que les besoins en capacité de stockage, bande passante réseau et personnel informatique.
- ☒ **Amélioration des niveaux de service de sauvegarde et de restauration** : la déduplication améliore nettement les performances de sauvegarde et permet ainsi de respecter les fenêtres de sauvegarde plus courtes. Elle optimise en outre les performances de restauration par rapport aux méthodes à accès séquentiel (bande) grâce au stockage sur disque à accès aléatoire.
- ☒ **Sauvegarde sur disque plus économique que la sauvegarde sur bande** : la déduplication facilite l'utilisation de la sauvegarde sur disque pour un plus large éventail d'applications. La sauvegarde sur bande est encore privilégiée dans les datacenters en raison de ses avantages économiques et de ses caractéristiques d'archivage. Toutefois, la déduplication entraînant une diminution du rapport coût de stockage/Go, la sauvegarde sur disque est appelée à devenir aussi économique, si ce n'est plus, que la sauvegarde sur bande.
- ☒ **Réduction de l'empreinte carbone** : dans la mesure où la déduplication permet de réduire les besoins en alimentation, en ventilation et en espace pour le stockage, elle diminue également l'empreinte carbone et facilite l'écoresponsabilité des entreprises.

La technologie de déduplication répond à un grand nombre des défis auxquels sont confrontées les petites comme les grandes entreprises en matière de sauvegarde depuis plus d'une dizaine d'années, en particulier l'explosion du volume de données, le raccourcissement des fenêtres de sauvegarde, la nécessité de réduire les délais de restauration en cas de panne d'exploitation ou de sinistre, etc.

Le tableau 1 décrit la grande diversité des défis actuels liés à la sauvegarde et les solutions apportées par la déduplication.

TABLEAU 1 :

Défis liés à la sauvegarde et impact de la déduplication

Défis liés à la sauvegarde	Impact de la déduplication
Les fenêtres de sauvegarde sont raccourcies car les systèmes sont opérationnels 24x7 pour satisfaire les besoins des clients dans le monde.	Les sauvegardes traditionnelles impliquent le transfert de gros volumes de données redondantes, difficiles à gérer compte tenu de fenêtres de sauvegarde restreintes, voire inexistantes. Deux méthodes permettent d'augmenter la quantité de données qu'il est possible de sauvegarder dans la fenêtre disponible. D'une part, la déduplication alliée à un logiciel de sauvegarde réduit le volume de données à sauvegarder. D'autre part, les systèmes de stockage avec déduplication rapide en continu permettent d'accélérer les performances de la cible de sauvegarde.
Les délais de restauration sont raccourcis pour réduire le coût des périodes d'interruption.	La déduplication diminue les coûts de sauvegarde sur disque en réduisant le volume de données à sauvegarder. La sauvegarde des données sur disque plutôt que sur bande raccourcit nettement les délais de restauration pour un large éventail d'applications.
La fiabilité des sauvegardes n'élimine pas les risques liés aux restaurations de données.	L'utilisation de bandes à des fins de sauvegarde introduit des risques d'erreurs inhérentes aux supports (défaillance du support, salissure des têtes, etc.), d'épuisement du support disponible ou de défaillances matérielles. La déduplication ayant recours à des disques pour la protection des données, ces risques de défaillances sont fortement réduits, voire éliminés. Le recours à des disques facilite également les bilans de santé et d'autres mesures d'autoréparation ou de prévention des défaillances.
Le développement de la virtualisation de serveurs se traduit par une diminution des ressources disponibles pour la sauvegarde, ce qui peut augmenter les délais de sauvegarde et la charge des fenêtres de sauvegarde.	La déduplication évite le traitement par des ressources partagées des données redondantes, ce qui réduit les conflits d'accès aux ressources physiques et accélère les sauvegardes de machines virtuelles. Elle permet également une rétention plus longue des données de sauvegarde des machines virtuelles sur un espace de stockage nettement plus réduit, garantissant ainsi une reprise rapide des opérations à partir de disques et jamais à partir de bandes.
La croissance des données se traduit par l'impossibilité de sauvegarder toutes les données dans les fenêtres de sauvegarde disponibles.	Les entreprises enregistrent une croissance moyenne annuelle de 50 % de la quantité de données à protéger. Cette augmentation est difficilement compatible avec les seules fenêtres de sauvegarde nocturnes et les méthodes traditionnelles de sauvegarde. La déduplication permet de résoudre cette problématique et de sauvegarder efficacement des Datasets en pleine expansion.
La réalisation de copies hors site sécurisées à l'aide des méthodes sur bande traditionnelles expose les données à un risque de vol ou de perte.	Le recours aux bandes amovibles pour le stockage hors site en cas de sinistre introduit des risques inhérents à tout support physique. Alliée à des procédures de réplique sécurisées, la déduplication permet de conserver une copie électronique hors site, éliminant ainsi toute manipulation manuelle de bandes pour une sécurité renforcée.

TABLEAU 1 :

Défis liés à la sauvegarde et impact de la déduplication

Défis liés à la sauvegarde	Impact de la déduplication
Les données distribuées dans les succursales distantes nécessitent une protection et une restauration centralisées.	Pour une sauvegarde, une restauration et une gestion améliorées, les succursales distantes remplacent les processus autonomes de sauvegarde sur bande par une approche de sauvegarde centralisée, depuis les filiales vers le datacenter. Grâce à la déduplication, il est possible d'envoyer d'importants volumes de données de sauvegarde vers un datacenter centralisé, via des liaisons WAN encombrées.
Les coûts d'infrastructure de sauvegarde augmentent pour satisfaire les besoins en matière de capacité et de fenêtres de sauvegarde.	La plupart des entreprises pallient l'explosion des données et la restriction des fenêtres de sauvegarde en étendant leur infrastructure de bandes. L'ajout de lecteurs de bandes et l'automatisation peuvent apporter une réponse aux goulots d'étranglement et accélérer les sauvegardes, mais moyennant une augmentation des coûts et des charges de gestion. La déduplication résout le problème à la racine. En effet, elle met un frein à la spirale des dépenses en infrastructure de bandes et permet de satisfaire les besoins croissants en matière de capacité et de fenêtres de sauvegarde.

Source : IDC, 2010

DÉDUPLICATION : DÉFINITION ET CONDITIONS DE MISE EN OEUVRE

Définition de la déduplication

IDC définit la déduplication comme une technologie qui normalise les données dupliquées et les traite en tant qu'objet de données partagées unique afin d'optimiser la capacité de stockage. Plus précisément, la déduplication des données désigne un algorithme qui recherche les données dupliquées (par exemple, blocs, segments) et élimine les données en double ainsi localisées. Une fois détectées, les données dupliquées ne sont pas conservées. Un « pointeur de données » est modifié de sorte que le système de stockage renvoie à une copie exacte de l'objet de données déjà stocké sur le disque. En outre, la déduplication réduit les coûts liés à la conservation de plusieurs copies du même objet de données.

La déduplication des données est, le plus souvent, associée à des processus de comparaison au niveau des sous-fichiers. La technique de stockage en instance unique est différente en cela qu'elle compare les données au niveau du fichier ou de l'objet. La déduplication au niveau du sous-fichier examine un fichier et le scinde en segments. Ceux-ci sont ensuite évalués pour trouver si des données redondantes existent sur plusieurs systèmes et sites. La déduplication diffère par ailleurs de la compression, qui réduit la taille de l'objet et non l'espace qu'il occupe dans plusieurs fichiers ou parties de fichier. En outre, les données dupliquées peuvent également être compressées afin d'économiser davantage l'espace.

À quel niveau exécuter la déduplication

La déduplication des données à des fins de sauvegarde peut s'effectuer au niveau de la source ou de la cible. Par exemple, une déduplication à la source peut avoir pour effet de réduire la taille des données de sauvegarde au niveau du client (serveur Exchange ou serveur de fichiers), de sorte que seules les données uniques de sous-fichiers sont transférées vers le réseau local ou étendu au cours du processus de sauvegarde. Lorsque la déduplication est réalisée à la source, la technologie de déduplication est intégrée dans l'architecture de l'application de sauvegarde. La déduplication au niveau de la cible peut, par exemple, consister à réduire la taille des données de sauvegarde, une fois qu'elles ont été transmises via le réseau local et ont atteint un système de stockage avec déduplication. La déduplication à la source permet de réaliser des économies de bande passante LAN et WAN, de fenêtre de sauvegarde et de stockage. Lorsque la déduplication est réalisée sur la cible, la technologie de déduplication est intégrée dans le contrôleur de stockage du système de stockage lui-même. Cette approche permet de faire des économies de stockage. Elle est compatible avec les logiciels de sauvegarde existants et peut réduire la charge sur le réseau étendu de la réplication. Les avantages, tout comme les délais et les coûts de la déduplication, varient selon le mode choisi. Les entreprises doivent évaluer leurs problèmes de sauvegarde actuels et comparer les avantages respectifs de ces deux méthodes de déduplication.

Déduplication à la source

La déduplication à la source (logiciel de sauvegarde client) offre de nombreux avantages qui dépassent la simple optimisation de la capacité de stockage. Cette méthode entraîne une réduction significative de la quantité de données

transférée du périphérique source vers le référentiel de stockage, libérant ainsi la charge sur une infrastructure virtuelle/physique et les liaisons LAN/WAN saturées. Dans la mesure où seuls les segments de données de sous-fichiers nouveaux ou modifiés sont transférés du périphérique source vers le référentiel de stockage, la quantité de données transférée est considérablement réduite et permet ainsi de réaliser des sauvegardes quotidiennes complètes extrêmement rapides. L'augmentation incrémentielle de la charge sur le CPU client lors de la déduplication à la source peut atteindre 15 %, mais la sauvegarde est plus rapide par comparaison aux méthodes traditionnelles. En outre, certaines architectures offrent des mécanismes de régulation permettant de gérer toute augmentation de charge à court terme. L'impact global de la déduplication à la source est considérablement plus faible que celui des agents classiques sur une période de sept jours. La déduplication à la source offre également une grande flexibilité de déploiement, car les bureaux distants de petite taille peuvent simplement mettre en oeuvre un agent de sauvegarde logiciel. Pour les environnements caractérisés par des bases de données très volumineuses ou des taux de modification de données quotidiens élevés, il peut être préférable d'opter pour la déduplication sur la cible. Heureusement, les fournisseurs disposent généralement d'outils d'évaluation des données pour aider les clients à choisir la solution la mieux adaptée à leurs besoins.

Déduplication sur la cible

La déduplication sur la cible optimise la capacité de stockage des disques de sauvegarde, car seules les données de sous-fichiers uniques et nouvelles sont stockées sur disque. Toutes les données de sauvegarde sont envoyées à la cible de déduplication à l'aide d'un logiciel de sauvegarde traditionnel. L'intégration dans votre infrastructure existante est ainsi transparente. Cette méthode n'allège la charge sur les fenêtres de sauvegarde disponibles que si la cible de sauvegarde précédente, généralement les bandes, constituait un goulot d'étranglement au niveau des performances des solutions de sauvegarde. Dans la déduplication sur la cible, le système de stockage lui-même (également appelé système de stockage avec déduplication) effectue la déduplication. La protection des données et la reprise après sinistre sont ainsi optimisées et les serveurs d'applications sont délestés du processus de déduplication. La déduplication sur la cible est facile à mettre en oeuvre par la création d'un système de stockage rapide et indépendant des applications (pouvant être connecté en tant que stockage rattaché au réseau [NAS] via Ethernet ou librairie de bandes virtuelles [VTL] sur Fibre Channel). Aucun logiciel client ni aucune autre configuration ne sont requis. Les systèmes de stockage avec déduplication sont souvent utilisés avec des bases de données et des Datasets volumineux. La déduplication sur la cible peut être effectuée à la fois dans les datacenters centraux dans le cas de volumes de données élevés et sur les succursales distantes dans le cas d'une sauvegarde locale suivie par une réplication vers un datacenter central.

Quand exécuter la déduplication

Il existe aujourd'hui deux approches différentes déterminant à *quel moment* exécuter le processus de déduplication : la déduplication en continu ou en post-traitement. La déduplication en continu élimine les données redondantes avant leur écriture sur le disque, rendant superflue toute zone de stockage temporaire sur le disque. La déduplication post-traitement analyse et réduit la quantité de données après leur stockage sur disque. Elle requiert donc une zone de stockage temporaire disposant

d'une capacité complète pour démarrer le processus de déduplication. Lors du choix de l'approche, les entreprises doivent prendre en considération des facteurs tels que la vitesse de sauvegarde et la capacité des disques.

La méthode de la déduplication en continu se révèle efficace et économique plus rapidement. Elle réduit la capacité brute des disques requise dans le système, car le Dataset complet et non encore dédupliqué n'est jamais écrit sur disque. Si la réplication est prise en charge en tant que partie intégrante du processus de déduplication en continu, ce dernier optimise également le temps de restauration, car le système ne doit pas attendre d'avoir absorbé l'intégralité du Dataset, puis de l'avoir dédupliqué avant de pouvoir commencer la réplication vers le site distant.

Le processus de déduplication post-traitement ne peut commencer avant le transfert complet des données sur disque. Cette approche requiert une capacité initiale plus élevée que les solutions en continu. En outre, la méthode post-traitement est assortie d'un délai avant la fin de la déduplication, ainsi qu'au cours de la réplication. Il existe également un risque d'incohérence entre un système local et un système distant en raison de la présence de deux emplacements de stockage, chacun avec ses règles et ses comportements à gérer.

Comment exécuter la déduplication

Les modalités d'exécution de la déduplication dépendent de l'implémentation. La méthode de déduplication basée sur le hachage scinde un fichier ou un flux de sauvegarde en segments de données de sous-fichiers de taille fixe ou variable. Une valeur de hachage est calculée pour chaque segment. Ce processus attribue un numéro unique à chaque segment, qui est ensuite stocké dans un index. Si un fichier est modifié, seules les données de sous-fichiers modifiées sont sauvegardées, les modifications ne nécessitant pas la sauvegarde d'un fichier entièrement nouveau. Dans le cas d'implémentations basées sur le hachage, il est important de déterminer si la taille des segments doit être fixe ou variable. La méthode à longueur variable peut, selon le type de contenu, ajuster dynamiquement la taille des segments en fonction des segments de données redondantes dont la position a été déplacée ou décalée dans un flux d'octets lors d'une modification de fichier. La méthode à longueur fixe ne reconnaît pas les données redondantes qui ont été déplacées ou décalées, de sorte qu'elle sauvegarde à nouveau des segments semblant uniques, même s'ils sont déjà présents dans le référentiel de sauvegarde. L'index de hachage est généralement conservé en mémoire, mais, à mesure que sa taille augmente, il peut déborder de la mémoire sur le disque et requérir des E/S de disque pour la recherche des valeurs de hachage. Les fournisseurs proposent des réponses diverses à ces défis technologiques, offrant des résultats allant d'une résolution du problème à une nette dégradation des performances.

Autre méthode utilisée : la déduplication des données basée sur le delta (également connue sous le nom de différenciation ou d'encodage delta), qui stocke ou transmet les données sous forme de différences par rapport à une copie de base. La copie de base est une copie complète des données utilisée pour recréer d'autres versions des données. La déduplication des données basée sur le delta peut être effectuée au niveau des blocs ou des octets. Plutôt que d'utiliser un hachage pour déterminer les données nouvelles sur le réseau, une méthode de différenciation delta analyse et indexe un flux de données entrant, en vue de rechercher des données similaires à celles déjà stockées. L'approche basée sur le delta présente des avantages, notamment la baisse du taux

d'utilisation du CPU, car il n'est pas nécessaire de calculer un hachage fort. En revanche, un processus de différenciation delta requiert un grand nombre d'E/S de disque pour comparer les données anciennes avec les nouvelles données entrantes. Par conséquent, l'avantage à long terme de chaque approche peut dépendre des améliorations des performances au niveau du CPU, comparé à celles offertes par la technologie de disque.

Le taux de déduplication dépend également de la capacité du moteur de déduplication à reconnaître des marqueurs insérés dans le flux de données par l'application de sauvegarde, ou un format de données spécifique (comme une application de sauvegarde, des données Microsoft Exchange, etc.). Pour repérer les marqueurs et le format de données, il est nécessaire de déterminer l'endroit où les métadonnées spécifiques de l'application sont injectées dans un flux. Si le moteur de déduplication peut déterminer les décalages de marqueurs ou le format de données, il peut adapter la taille des segments de manière à ce qu'elle soit idéale pour le format de données en fonction de leur application native, et fournir des résultats de déduplication potentiellement supérieurs. Toutefois, l'utilisation de cette approche nécessite d'acquérir et d'entretenir une compréhension des formats changeants de chaque application de sauvegarde (NetWorker, NBU, TSM, etc.) et utilisateur (Oracle, Exchange, etc.).

FACTEURS À PRENDRE EN COMPTE LORS DE L'ÉVALUATION DE LA TECHNOLOGIE DE DÉDUPLICATION

Plusieurs types de produit dotés de fonctions de déduplication sont disponibles dans le commerce. Applications de sauvegarde, appliances, bibliothèques de bandes virtuelles, solutions d'optimisation WAN et sous-systèmes de stockage sur disque principal peuvent tous comporter une fonction de déduplication sous une forme ou une autre. Il est important pour une entreprise de déterminer les problèmes à résoudre par application ou par type de données avant d'opter pour un type de déduplication. Chaque méthode de déduplication présente des avantages différents en termes de capacité, de performances et d'efficacité réseau.

- 1. Taux de déduplication :** le taux de déduplication obtenu dépend d'une multitude de facteurs, comme le type de données, le taux de modification des données, les périodes de rétention, les segments de longueur variable ou fixe, les règles de sauvegarde, la détection des formats de fichier, etc. Les études menées par IDC en environnements réels font apparaître des taux de déduplication de stockage sur disque back-end compris entre 8:1 et 22:1 en fonction des facteurs susmentionnés. Par rapport aux méthodes traditionnelles de sauvegarde complète, les solutions de déduplication à la source peuvent réduire sensiblement la consommation quotidienne de bande passante réseau. Toutefois, comme pour toutes les mesures de performance, le résultat obtenu varie en fonction de l'environnement. Les entreprises doivent examiner avec soin les garanties de rendement, d'évolutivité ou de performances offertes, et tester la déduplication sur leurs sites avec leurs propres Datasets.
- 2. Rôle de la compression, du chiffrement et du multiplexage :** codage des données destiné à réduire leur taille de stockage, la compression peut constituer un excellent complément à la déduplication. La compression est optimisée pour un seul objet et réduit sa taille, tandis que la déduplication

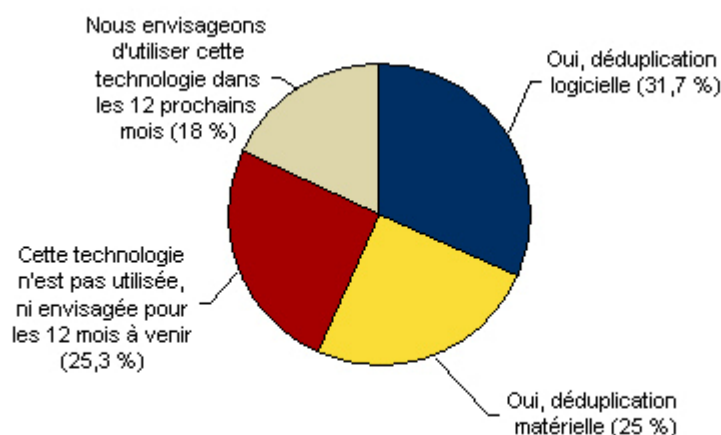
s'applique à tous les objets. Toutefois, la compression peut être appliquée à des données déjà dédupliquées afin d'offrir des économies d'espace supplémentaires. Si la déduplication est appliquée à un fichier déjà compressé (ou chiffré), le bénéfice de la déduplication sera néanmoins négligeable, voire nul, à moins que ce fichier compressé ne soit sauvegardé à nouveau. Les entreprises peuvent tirer d'autres avantages de la déduplication associée à la compression, si elles procèdent à la déduplication des données en premier lieu. Il est également nécessaire de tenir compte de l'utilisation courante du multiplexage pour les sauvegardes, processus qui imbrique les données de plusieurs clients dans un flux unique transféré sur un lecteur de bande. Ce processus complique la détection des segments de données existants. Pour tirer parti de la déduplication, les entreprises doivent désactiver le multiplexage, une fonction offerte par la plupart des applications de sauvegarde permettant d'éviter le défaut de synchronisation entre le déroulement de la bande et le flux de données (*shoe-shining*) et d'améliorer les performances lors de l'écriture des données sur un support de bande. Toutefois, dans ce scénario, la désactivation du multiplexage ne réduit pas à néant les améliorations permises en termes de performances de la sauvegarde.

3. **Déduplication pour les machines virtuelles** : l'utilisation de machines virtuelles en production renforce les besoins de protection et de restauration de la machine virtuelle, de l'hôte physique et des fichiers. Parmi les options de sauvegarde des machines virtuelles figurent la sauvegarde au niveau image et/ou fichier à l'aide d'un agent de sauvegarde s'exécutant dans un système invité ou dans une console de service, ou l'utilisation de l'API de sauvegarde d'un fournisseur de solutions de virtualisation. Les solutions de sauvegarde traditionnelles sont inefficaces pour la sauvegarde des machines virtuelles, car elles déplacent des quantités importantes de données redondantes et requièrent un grand nombre de cycles de CPU pour effectuer une sauvegarde, entraînant une baisse des performances de sauvegarde et un taux limité de consolidation des serveurs. La déduplication permet de pallier ces limites. Avec la déduplication à la source, les données dupliquées ne transitent jamais par l'infrastructure physique sous-jacente partagée. Les sauvegardes quotidiennes complètes sont donc rapides et efficaces. La déduplication peut également être effectuée globalement sur les fichiers VMDK afin d'éliminer la sauvegarde des données redondantes sur l'ensemble des systèmes virtuels. L'utilisateur est ainsi en mesure de restaurer des machines virtuelles sans bande et d'effectuer la reprise après sinistre à l'aide des fonctions de réplication des systèmes de déduplication de manière plus performante et efficace.
4. **Déduplication pour les succursales distantes** : tout comme les datacenters, les succursales distantes requièrent à la fois une restauration locale et une reprise après sinistre (à distance). Les succursales distantes posent toutefois des défis particuliers. Elles se caractérisent généralement par une bande passante de réseau étendu limitée, l'absence de personnel informatique dédié et un nombre disproportionné de sites connectés aux datacenters régionaux ou principaux. La déduplication permet de réduire au minimum les transferts de données sur le réseau étendu et d'éliminer les données redondantes sur les sites des succursales et dans le datacenter. Compte tenu du faible effectif de personnel informatique sur les sites des succursales distantes, de nombreuses entreprises cherchent à réduire

l'encombrement du matériel de stockage sur les sites distribués. La déduplication à la source peut être déployée par voie logicielle, ce qui résout le problème. Des systèmes de déduplication plus petits peuvent également être déployés sur les sites distants lorsqu'une restauration rapide en local est requise et certains fournisseurs prennent en charge la réplication depuis ces systèmes vers un datacenter central. Une étude récente menée par IDC a examiné l'utilisation de la déduplication dans les succursales distantes et le type déployé (voir la figure 1).

FIGURE 1

Utilisation de la technologie de déduplication pour la protection des données des succursales distantes



n = 300

Source : IDC, Remote Branch Special Study, 2009

- Déduplication pour les datacenters de production/reprise après sinistre** : Les grands datacenters ont toujours des difficultés à respecter leur fenêtre de sauvegarde pour une partie au moins de leurs applications et ne peuvent en aucun cas sacrifier les performances de sauvegarde. Cette situation peut plaider en faveur d'une méthode de déduplication incluant la déduplication à la source et sur la cible en fonction de l'application et de l'environnement. L'optimisation de la bande passante réseau dans un datacenter peut être moins prioritaire que la réplication à distance sur un site de reprise. Cependant, du fait du raccourcissement continu des fenêtres de sauvegarde, la largeur de bande passante réseau est appelée à revêtir une importance cruciale.
- Déduplication et réplication** : la réplication est le prochain défi à relever en matière de déduplication. Des fournisseurs établis ont fait la preuve de son efficacité lorsqu'elle est correctement conçue et mise en oeuvre, et les utilisateurs qui ont pu évaluer cette technologie se sont montrés enthousiastes et très demandeurs. La déduplication est déployée dans les environnements d'entreprise, sur les sites principaux et secondaires,

permettant de gagner en efficacité tout en réduisant les coûts d'infrastructure. L'utilisation de la réplication à distance revêt une importance majeure, dans un contexte où les entreprises limitent leur recours aux bandes dans les sites distants, mais doivent continuer à prendre en charge les bandes dans un site centralisé à des fins d'archivage et de conformité, notamment. Les besoins des utilisateurs en matière de réplication sont de plus en plus stricts et portent notamment sur les aspects suivants :

- ❑ **Réplication compatible avec la déduplication, qui réplique un Dataset déduplicqué et non un volume complet** : certains fournisseurs proposent des services de réplication avec un produit compatible avec la déduplication. Les entreprises doivent cependant s'assurer que la fonction de réplication est compatible avec la déduplication.
- ❑ **Réplication « tout ou rien » et réplication au niveau des répertoires/bandes** : dans certaines situations, une réplication du système complet est requise, tandis que, dans d'autres, il est important de pouvoir choisir les éléments partagés ou les bandes virtuelles à répliquer.
- ❑ **Surveillance de la réplication, réglage des performances et résolution des problèmes** : malgré la déduplication, dans la plupart des grandes entreprises, les volumes de données à répliquer restent importants. Ils sont gérés via un processus de réplication programmé ou asynchrone, le processus de réplication et la consommation de bande passante étant surveillés. Des outils de réglage et de résolution des problèmes permettent de s'assurer que le processus de réplication ne dévie pas de la fenêtre de réplication disponible.
- ❑ **Réplication programmée et en temps réel pour des liaisons à latence élevée et faible** : certaines liaisons ou certains sites requièrent une réplication en temps réel, tandis que d'autres peuvent se contenter d'une réplication programmée. Les caractéristiques des filiales distantes varient considérablement, certaines pouvant disposer de liaisons à faible latence alors que les liaisons entre deux datacenters peuvent ne pas présenter ce problème.

7. **Réplication à distance et migration** : bien que la déduplication permette de réduire les besoins en stockage et/ou la transmission de données redondantes, elle requiert l'établissement d'une base, ou sauvegarde, initiale. Dans le cas d'une déduplication et d'une réplication depuis les filiales vers le datacenter, les utilisateurs doivent réfléchir à la manière de créer cette base initiale via des liaisons présentant des contraintes de bande passante. La plupart des fournisseurs offrent un service de réplication à distance permettant de créer cette base initiale soit par l'intermédiaire d'un processus de réplication en masse compatible avec la déduplication avec des systèmes côte à côte, soit par l'utilisation d'une série de bandes provenant de la dernière sauvegarde complète et leur restauration locale dans un système de déduplication. Avec des cycles d'actualisation du stockage de trois à cinq ans, un autre facteur à prendre en compte est la manière dont la migration s'effectue ainsi que les éventuelles interruptions de service dans l'environnement existant.
8. **Sélection des fournisseurs** : les fournisseurs affirment beaucoup de choses sur leurs solutions de déduplication. L'étude menée par IDC montre cependant que les produits de déduplication généralement disponibles n'offrent pas tous les performances annoncées. Les entreprises doivent être attentives à l'ancienneté de mise sur le marché d'un produit de déduplication

donné, au nombre de clients utilisant le produit en production et à la maturité du produit dans les environnements réels. Elles doivent se renseigner précisément sur l'évolutivité du produit, notamment en demandant des références utilisateur et une matrice de support d'application et/ou de système. Les entreprises qui choisissent de ne pas conduire de validation fonctionnelle risquent d'avoir de mauvaises surprises en termes de performances et de fiabilité.

9. **Exemples d'utilisation de la déduplication** : la déduplication est une technologie appelée à remonter dans la pile d'infrastructure de stockage. À ce jour, cette technologie a été largement déployée dans le domaine de la sauvegarde qui se caractérise par la présence de grandes quantités de données redondantes. Ces mêmes données sont sauvegardées chaque semaine et monopolisent inutilement des ressources de serveur, de réseau et de stockage. Certaines entreprises commencent à se tourner vers la technologie de déduplication existante ou à la tester au sein d'environnements de stockage principal dans le cadre d'une approche NAS. Toutefois, cette implémentation requiert davantage de performances pour éviter tout problème de latence et de temps de réponse insuffisants. Aujourd'hui, la technologie de la déduplication convient parfaitement à la sauvegarde des environnements de machines virtuelles, de sites distants/succursales et de datacenters.

GAMME DE SOLUTIONS DE DÉDUPLICATION EMC

EMC offre une large gamme de produits et de services de sauvegarde et de restauration pour aider les clients à réduire leurs coûts informatiques et à augmenter l'efficacité de la sauvegarde. Les solutions de sauvegarde basées sur la déduplication comprennent le logiciel de sauvegarde avec déduplication EMC Avamar, les systèmes de stockage avec déduplication EMC Data Domain, ainsi qu'EMC NetWorker, qui peuvent être déployés avec Avamar, Data Domain et d'autres systèmes cibles tiers. De plus, bien que cela dépasse le cadre de ce document, EMC propose une solution de déduplication pour le stockage principal et les données de sauvegarde avec son système NAS EMC Celerra, ainsi qu'une solution de déduplication des archives sur disque avec sa gamme de produits Centera.

EMC Avamar

Les logiciels de sauvegarde avec déduplication EMC Avamar intègrent une technologie de déduplication pour identifier les données redondantes au niveau de la source et réduire ainsi au minimum le volume de données de sauvegarde avant leur envoi via le réseau LAN/WAN. Grâce à Avamar, une entreprise bénéficie d'une diminution du volume des données et de sauvegardes quotidiennes complètes et rapides pour les environnements VMware, les bureaux distants, les ordinateurs de bureau/portables et les serveurs LAN et NAS. Avamar procède également à une déduplication régulière des données de sauvegarde sur l'ensemble des sites et des serveurs. Contrairement aux produits reposant sur des méthodes de restauration traditionnelles, Avamar peut restaurer rapidement des données en une seule étape, vous épargnant ainsi la restauration de la dernière sauvegarde complète appropriée et des sauvegardes incrémentielles suivantes pour parvenir au point de

restauration voulu. Les fonctions d'Avamar se démarquent radicalement des applications de sauvegarde classiques.

L'agent Avamar détecte les nouveaux fichiers et ceux qui ont été modifiés. Il n'a pas besoin de parcourir la totalité de l'arborescence du système de fichiers pour identifier les données nouvelles ou modifiées. Il commence en effet par vérifier le cache de fichiers local pour localiser ces fichiers. Une fois les fichiers identifiés, l'agent scinde les fichiers nouveaux ou modifiés en segments de données de sous-fichiers de longueur variable et attribue une valeur de hachage (ID unique) à chaque segment. L'agent vérifie ensuite dans le cache de hachage local si le hachage existe et a déjà été sauvegardé. Si tel est le cas, il ne le sauvegarde pas de nouveau. Enfin, l'agent communique avec le serveur Avamar pour déterminer si la valeur de hachage est unique ou existe déjà. Si le segment de données est nouveau, il est envoyé via le LAN/WAN lors de la sauvegarde complète quotidienne.

Ces processus entraînent une augmentation du taux d'utilisation du CPU de l'hôte par rapport à un agent de sauvegarde traditionnel. Toutefois, la sauvegarde assurant une protection efficace des seuls segments de données nouveaux sur le réseau, les sauvegardes Avamar sont nettement plus rapides que les sauvegardes complètes et incrémentielles classiques. Par exemple, une sauvegarde incrémentielle qui durait auparavant près de 10 heures ne prendra désormais pas plus d'une heure avec Avamar, ce qui réduit de 50 à 5 heures le temps consacré aux sauvegardes incrémentielles hebdomadaires. Les sauvegardes complètes quotidiennes réalisées avec Avamar sont également plus rapides que les sauvegardes complètes classiques.

En outre, Avamar offre désormais une protection des ordinateurs de bureau/portables qui s'exécute en arrière-plan via des liaisons réseau existantes et ne fait pas perdre de précieux cycles de CPU. La sauvegarde des données des utilisateurs se fait automatiquement à la connexion de ces derniers, lors de fenêtres de sauvegarde normales, ou peut être effectuée par les utilisateurs, ce qui leur permet de restaurer leurs données à volonté.

Avec leurs fonctions de déduplication globale des données à la source, les solutions de sauvegarde et de restauration Avamar répondent aux besoins suivants :

- déploiement de machines virtuelles et évaluation d'une nouvelle stratégie de protection pour la restauration des serveurs physiques, des serveurs virtuels et des objets discrets ;
- amélioration des sauvegardes des succursales distantes pour bénéficier de sauvegardes complètes quotidiennes rapides, d'une gestion centralisée, d'une fiabilité accrue, d'une réplication sécurisée et d'un trafic de sauvegarde réduit sur les liaisons WAN encombrées ;
- limitation de la croissance des données, du raccourcissement des fenêtres de sauvegarde et du trafic réseau pour la sauvegarde des environnements NAS et de serveur de fichiers locaux ;
- protection des données de valeur, contenues dans les ordinateurs de bureau/portables utilisés dans les bureaux et par les professionnels itinérants.

Les configurations de déploiement d'EMC Avamar sont les suivantes :

Options clients

- Les agents logiciels Avamar sont déployés sur les systèmes à protéger (clients) sans matériel local supplémentaire (serveurs de média, par exemple). Ils sont disponibles pour la plupart des systèmes d'exploitation et les principales applications et bases de données. Contrairement à de nombreuses applications de sauvegarde, Avamar ne facture pas ses logiciels clients, mais fonctionne plutôt selon un système de licences basé sur la capacité dédupliquée. Les mises en oeuvre peuvent se révéler extrêmement économiques et être suivies d'un développement de la clientèle.

Options serveur (référentiel de sauvegarde)

- Serveurs Avamar tiers : le logiciel Avamar peut être acheté et déployé sur une vaste gamme de serveurs standard certifiés disposant d'une capacité de stockage sur disque interne.
- Avamar Data Store : cette solution tout en un évolutive inclut des composants logiciels Avamar préinstallés et préconfigurés sur un matériel EMC pour une commande, un déploiement et une maintenance simplifiés.
- Avamar Virtual Edition for VMware : cette configuration haut de gamme permet de déployer un serveur Avamar en tant qu'appliance virtuelle sur un hôte VMware ESX existant, en tirant parti des ressources et de la capacité de stockage sur disque associées.

Avamar se démarque des autres approches de déduplication à la source disponibles sur le marché. Par exemple, la technologie de déduplication d'Avamar, qui est basée sur l'utilisation de segments de données de sous-fichiers de longueur variable, offre une efficacité et des performances supérieures. Avamar s'appuie sur une architecture distribuée pour des performances et une capacité évolutives, chaque noeud incrémentiel augmentant les ressources de CPU, de mémoire, d'E/S et de stockage pour l'ensemble du système.

La grille Avamar utilise la technologie RAIN (Redundant Array of Independent Nodes) qui fournit une tolérance aux pannes et une haute disponibilité intégrées sur l'ensemble de la grille, et élimine les points unitaires de panne. Avamar répartit son index interne sur l'ensemble des noeuds Avamar pour une fiabilité, un équilibrage de la charge et une évolutivité optimaux. De plus, la solution Avamar vérifie quotidiennement et automatiquement que les données de sauvegarde peuvent être complètement restaurées, et le serveur Avamar exécute automatiquement un autotest deux fois par jour pour garantir son intégrité. Enfin, Avamar prend en charge un large éventail d'applications et de clients, notamment Exchange, SQL, Oracle, DB2, SharePoint, Lotus Notes et NDMP.

Avamar offre différents modes de protection des machines virtuelles et physiques. Les options de sauvegarde Avamar pour les environnements de machines virtuelles VMware incluent notamment :

- Agent Avamar dans le système d'exploitation invité** : l'agent Avamar installé dans chaque système d'exploitation invité offre une méthode de sauvegarde nettement plus efficace que les solutions de sauvegarde par

agent traditionnelles. Les agents Avamar simplifiés réduisent le volume de données de sauvegarde au niveau de l'invité, ainsi que les conflits d'accès et les besoins réseau pour les ressources partagées de CPU, de carte réseau, de disque et de mémoire. Dans la mesure où il ne sauvegarde que les données de sous-fichiers nouvelles ou uniques, Avamar permet d'effectuer rapidement des sauvegardes complètes quotidiennes.

- ☒ **Avamar pour la sauvegarde VCB ou de l'API vStorage** : un agent Avamar s'exécutant sur le serveur proxy ne sauvegarde que les données uniques et allège la charge de traitement des machines invitées. La déduplication couvre l'ensemble des fichiers VMDK et prend en charge la sauvegarde au niveau des images et des fichiers. La réplication Avamar permet le transfert rapide de fichiers VMDK sur le WAN dans des situations de reprise après sinistre.
- ☒ **Agent Avamar sur console ESX** : un agent Avamar installé sur la console ESX peut dédupliquer les données au niveau de l'ensemble des fichiers VMDK. Cette méthode fournit une option de sauvegarde et de restauration au niveau image, sans aucune dépendance vis-à-vis de VMware VCB ni des ressources de stockage partagées. Toutefois, elle ne fournit pas de restauration au niveau fichier.

EMC Data Domain

Les systèmes de stockage avec déduplication Data Domain réduisent considérablement la capacité de stockage sur disque nécessaire à la rétention et à la protection des données d'entreprise. En identifiant les données et les fichiers redondants à mesure de leur stockage, les systèmes Data Domain offrent un encombrement de stockage en moyenne 10 à 30 fois inférieur à celui du Dataset d'origine. Les données de sauvegarde peuvent être efficacement répliquées et récupérées via des réseaux existants, pour une reprise après sinistre rationalisée et des opérations sur bandes consolidées.

Les systèmes Data Domain sont disponibles dans plusieurs configurations qui diffèrent les unes des autres en termes de performances et de capacités. Plusieurs options logicielles améliorent les fonctions et la valeur de ces systèmes. Exemples :

- ☒ La famille Data Domain Appliance offre des systèmes de stockage avec déduplication à haut débit, économiques et évolutifs avec stockage intégré.
- ☒ La famille Data Domain DDX Array offre un système de stockage sur baie de disques évolutif et hautes performances, avec jusqu'à 56,7 pétaoctets de capacité logique et jusqu'à 86,4 téraoctets/heure de débit. Cette solution permet de gérer et de prendre en charge jusqu'à 2 880 sites distants lors de l'utilisation de contrôleurs DD880.
- ☒ La famille Data Domain Gateway propose des passerelles d'entreprise aux multiples avantages en termes de déduplication et de compression, pour les datacenters utilisant des systèmes de stockage externes tiers.
- ☒ Data Domain Replicator est une solution logicielle de réplication automatisée et efficace en réseau conçue pour la reprise après sinistre, la protection des données des bureaux distants et la consolidation de bandes multisite.

- ☒ Le logiciel Data Domain Virtual Tape Library émule plusieurs bibliothèques de bandes via une interface Fibre Channel, en vue de fournir un stockage avec déduplication pour les environnements SAN.
- ☒ Le logiciel Data Domain OpenStorage offre une intégration transparente entre les systèmes de stockage avec déduplication Data Domain et Symantec Veritas NetBackup.
- ☒ Grâce au logiciel Data Domain Retention Lock, les utilisateurs peuvent facilement mettre en oeuvre la déduplication avec verrouillage de fichiers afin de respecter les règles de gouvernance informatique et de conformité. Les administrateurs informatiques disposent de la flexibilité opérationnelle nécessaire pour gérer efficacement l'entreprise au quotidien, au meilleur coût. Ce logiciel inclut des fonctions sécurisées de destruction (shredding) des données.

Les solutions de déduplication Data Domain offrent un niveau de stockage consolidé des données de sauvegarde, nearline et d'archivage. En outre, les solutions Data Domain s'intègrent dans l'infrastructure existante et offrent un degré satisfaisant de protection des investissements, ce qui les rend idéales pour les environnements suivants :

- ☒ datacenters affichant une croissance exponentielle des données à sauvegarder et des données à contenu fixe et devant faire face à des exigences de plus en plus fortes en matière de restauration opérationnelle et de reprise après sinistre ;
- ☒ entreprises de grande taille devant prendre en charge la consolidation depuis des bureaux distants vers le datacenter, pour la protection des données et la reprise après sinistre ;
- ☒ infrastructure de sauvegarde et d'archivage nécessitant une cible de déduplication consolidée ;
- ☒ infrastructures fondées sur des systèmes et des supports de bandes et devant revoir la conception de la sauvegarde et de la restauration pour résoudre d'importants problèmes de coûts et de gestion.

L'architecture d'invulnérabilité des données Data Domain garantit l'intégrité de l'ensemble des données de sauvegarde, en offrant des niveaux élevés de protection, de vérification des données et d'autoréparation. Les principales fonctions de protection de l'intégrité des données sont les suivantes :

- ☒ vérification continue de la reprise, afin de garantir l'intégrité des données de sauvegarde et leur restauration depuis tous les niveaux du système, tout au long de leur cycle de vie ;
- ☒ vérification de l'unicité assurant une protection contre les collisions de hachage aléatoires et malveillantes afin de garantir la réussite des restaurations de données ;
- ☒ RAID à double parité sur deux disques (RAID 6) afin de protéger le système d'un maximum de deux pannes disque simultanées.

Des niveaux supplémentaires de protection de l'intégrité des données sont intégrés dans l'architecture d'invulnérabilité des données pour détecter les pannes et les corriger, afin que la restauration des données de sauvegarde ne présente aucun risque.

Le logiciel Data Domain Replicator transfère uniquement les modifications uniques déduplicées et compressées via n'importe quel réseau IP. Ses besoins en bande passante et le temps et le coût requis sont donc bien inférieurs à ceux des méthodes classiques de réplication. Si le système de destination de la réplication est le même pour plusieurs systèmes Data Domain, l'effet de la déduplication gagne en efficacité, car le système de destination ne stocke qu'une seule fois chaque segment unique sur tous les flux de réplication entrants, réduisant encore davantage la bande passante nécessaire.

Les principaux avantages de la technologie de réplication Data Domain sont les suivants :

- stockage distant sur des disques vault WAN pour la reprise après sinistre, qui offre une protection des données basée sur le réseau, en répliquant de façon sécurisée et automatique les données de sauvegarde vers un site distant sécurisé ;
- protection des données des bureaux distants, permettant le stockage distant des données de sauvegarde sur des disques vault, depuis de nombreuses succursales vers un hub ou datacenter central ;
- réplication en cascade, permettant de répliquer des données déduplicées depuis un bureau distant vers un datacenter central, puis vers d'autres sites tels qu'un site de reprise, pour une capacité de restauration et une protection contre les sinistres améliorées ;
- consolidation de bandes, rendant inutile la duplication des données de sauvegarde sur chaque site distant et permettant une infrastructure de bandes considérablement réduite, uniquement dans un hub central avec la présence d'une équipe informatique.

EMC NetWorker

EMC NetWorker est une application de sauvegarde d'entreprise qui centralise les opérations de sauvegarde et de restauration. NetWorker fournit une plate-forme commune prenant en charge de nombreuses options de protection des données, notamment la sauvegarde sur disque, la réplication, la protection continue des données et la déduplication dans des environnements physiques et virtuels. La polyvalence de NetWorker en fait le logiciel de sauvegarde idéal pour les clients qui choisissent de simplifier la gestion de l'ensemble de leurs environnements, des datacenters importants aux bureaux distants. La principale application NetWorker fournit une déduplication à la source grâce à son intégration avec la technologie de déduplication d'EMC Avamar et peut également tirer parti des solutions de déduplication sur la cible, telles que les systèmes Data Domain, dans le cadre des opérations qu'elle effectue. Les entreprises qui utilisent la déduplication NetWorker cherchent à :

- maîtriser la croissance des gros volumes de données dans les environnements NetWorker existants ;

- ☒ déployer une nouvelle stratégie de sauvegarde sur disque pour une restauration plus performante qui requiert encore l'usage de bandes physiques pour l'archivage ou les besoins à long terme ;
- ☒ répondre à un ensemble varié d'exigences, parmi lesquelles certaines requièrent une déduplication à la source et d'autres une déduplication sur la cible ;
- ☒ réduire les coûts et la complexité en regroupant plusieurs stratégies de protection des données dans une seule application.

La méthode de déduplication mise en oeuvre dans l'application NetWorker fut une initiative pionnière en termes d'intégration de la déduplication dans une application de sauvegarde classique. Le logiciel client NetWorker pour les sauvegardes non compatibles et compatibles avec la déduplication est un agent unique. Les fonctions de déduplication à la source ont été entièrement intégrées, réduisant ainsi au minimum les tâches de déploiement et de maintenance. La console NetWorker peut gérer et surveiller les deux types de sauvegarde : traditionnelle et avec déduplication. Les utilisateurs de NetWorker qui souhaitent bénéficier des avantages de la déduplication ne doivent supporter aucun coût supplémentaire.

Contrairement aux autres produits, NetWorker ne comporte aucune UGS ni tarif de logiciel supplémentaires pour l'intégration de la déduplication. Les utilisateurs de NetWorker peuvent ajouter à l'environnement de sauvegarde le moteur de déduplication approprié : le logiciel Avamar ou la solution back-end Data Domain. L'un des avantages de la déduplication compatible NetWorker est la prise en charge des bandes physiques, qui permet aux clients de continuer à utiliser des bandes dans la même application. Elle offre également la possibilité de tirer parti de la prise en charge des applications riches de NetWorker, en vue d'effectuer une reprise après sinistre et granulaire et de gérer des snapshots pour la sauvegarde hors hôte. NetWorker offre aux entreprises de puissantes fonctions de déduplication sans perturber leur environnement de sauvegarde actuel.

QUELLE APPROCHE ADOPTER POUR RELEVER LES DÉFIS ?

Comme le montre ce livre blanc, les différentes technologies et méthodes de déduplication offrent des avantages différents selon le contexte d'utilisation. C'est pourquoi il est important de savoir quel produit EMC convient le mieux à tel ou tel environnement. Le tableau 2 constitue un mémento permettant aux entreprises d'identifier la solution de sauvegarde et de restauration EMC la mieux adaptée à leurs besoins.

EMC propose une large gamme de produits avec des fonctions de déduplication. La déduplication n'étant pas un produit autonome, EMC doit enseigner aux clients la meilleure façon de tirer parti de cette technologie en fonction des problématiques propres à leur environnement. Cette sensibilisation, appuyée par des études de cas et des tests comparatifs d'évolutivité et de performances, permettra d'accroître la confiance des clients dans l'efficacité de la technologie d'un produit donné.

TABLEAU 2 :

Sélection d'un produit de déduplication EMC

	EMC NetWorker	EMC Data Domain	EMC Avamar
Déduplication pour la sauvegarde	<ul style="list-style-type: none">• Source• Déduplication en continu	<ul style="list-style-type: none">• Cible• Déduplication en continu	<ul style="list-style-type: none">• Source• Déduplication en continu
Idéale pour les environnements possédant les caractéristiques suivantes :	<ul style="list-style-type: none">• Environnements NetWorker• Prise en charge des bandes physiques requise• Environnements hétérogènes de grande taille	<ul style="list-style-type: none">• Nécessité de sauvegardes et restaurations ultrarapides• Réplication pour sauvegarde hors site• Prise en charge de l'environnement de sauvegarde actuel, sans modifications opérationnelles• Prise en charge du datacenter et des sites distants	<ul style="list-style-type: none">• Environnements virtuels• Succursales distantes• Serveurs LAN/NAS• Ordinateurs de bureau/portables
Options de déploiement	<ul style="list-style-type: none">• Agent NetWorker unique• Voir options de déploiement Avamar	<ul style="list-style-type: none">• Passerelle ou matériel d'appliance	<ul style="list-style-type: none">• Agent uniquement : pour bureaux distants de petite taille• Avamar Data Store : solution complète clé en main (matériel et logiciel)• Serveur tiers : création de son propre serveur Avamar• Avamar Virtual Edition : appliance virtuelle utilisant VMware ESX et le disque existants

Source : IDC, 2010

CONCLUSION

La technologie de déduplication accélère les sauvegardes et réduit les coûts informatiques. Les entreprises déploient différents types de solution avec déduplication pour faire face aux nombreux défis économiques et opérationnels que pose la croissance des volumes de données de sauvegarde. IDC considère la déduplication comme une fonction essentielle et incontournable que doivent intégrer les solutions de stockage pour résoudre ce problème. EMC est un fournisseur bien positionné sur ce marché, grâce à son offre de solutions adaptées à une grande variété d'environnements et de situations, qui permettent de répondre aux besoins des clients en matière de déduplication pour au moins les cinq ans à venir.

Note de copyright

Publication externe d'IDC Information and Data : l'utilisation de toute information IDC dans des publicités, des communiqués de presse ou des documents promotionnels est soumise à l'autorisation écrite préalable du vice-président ou du responsable régional d'IDC. Une telle demande doit être accompagnée d'un projet du document proposé. IDC se réserve le droit de refuser une utilisation externe pour tout motif.

Copyright 2010 IDC. Toute reproduction sans autorisation écrite est strictement interdite.