


5 astuces pour moderniser la déduplication des datacenters nouvelle génération (Tribune)

Les datacenters évoluent : les services informatiques doivent faire face aux exigences nouvelles de la virtualisation et du Cloud, ainsi qu'à l'émergence de technologies de stockage basées sur des workflows qui supplantent les solutions traditionnelles de protection des données. Ces tendances interagissent entre elles et rendent la gestion des données plus complexe, malgré les promesses de rationalisation des opérations, des ressources et d'automatisation de la prise de décisions. 

Les datacenters deviennent plus dynamiques et difficiles à gérer, les workflows de données sont de moins en moins linéaires, les entreprises sont par conséquent confrontées à une nouvelle réalité : elles doivent moderniser leurs stratégies en matière de déduplication, notamment en repensant leurs attentes quant aux appliances de sauvegarde dédiées.

Déduplication et datacenters de nouvelle génération : les nouvelles tendances

Un certain nombre de tendances conditionnent la modernisation de la déduplication au sein des entreprises. Voici les trois les plus répandus. Tout d'abord, les entreprises s'appuient de plus en plus sur des architectures Cloud variées et se doivent de sauvegarder vers ces nouveaux environnements de façon efficace. Ensuite, les services informatiques gèrent désormais des volumes toujours plus importants de données non structurées, ainsi que les métadonnées associées, et doivent pouvoir faire évoluer leurs ressources de sauvegarde en fonction de ce paramètre, tout en distinguant les données « déduplicables » de celles qui ne le sont pas. Enfin, la hiérarchisation efficace des données à sauvegarder et archiver pour un stockage et une restitution souples est de plus en plus essentielle aux opérations quotidiennes des entreprises.

Les entreprises, sous l'effet de l'ensemble de ces tendances, s'appuient donc sur un stockage dédié pour gérer la nature complexe de leurs défis métier spécifiques au secteur dans lequel elles évoluent, notamment les workflows qui régissent le transfert de leurs données. Il se peut que certains n'apprécient pas la déduplication à sa juste valeur, mais celle-ci rend la reprise après incident viable, même pour les mises en œuvre modestes. La déduplication jouera un rôle toujours plus important au sein des datacenters nouvelle génération. Les solutions présentant les frais d'exploitation les plus faibles auront donc le plus de chances de convaincre.

Étude des workflows de déduplication

Plusieurs facteurs doivent être pris en compte pour déterminer comment la déduplication peut être intégrée aux workflows et datacenters modernes. Il n'existe cependant pas de solution miracle pour maîtriser la complexité des centres de données. Les besoins en termes de type de données, contenu et fréquence d'accès doivent être évalués afin d'identifier la solution de déduplication la plus adaptée.

Les machines virtuelles, par exemple, requièrent de nombreuses applications de sauvegarde pour

fonctionner au sein de workflows virtuels plus dynamiques, qu'elles ne sont pas vraiment en mesure de gérer correctement. La gestion de ce type de données doit être différente de l'approche traditionnelle.

La gestion du volume croissant de données non structurées devient au même titre un aspect essentiel à toute stratégie de protection des données, qu'il s'agisse de l'archivage de contenu vidéo pour une remonétisation ultérieure, du transfert de données statiques à partir d'un stockage primaire ou de l'élaboration d'une infrastructure Cloud privée. En raison des contraintes d'évolutivité et d'accessibilité qui s'appliquent au stockage de ces données, la sauvegarde traditionnelle n'est tout simplement plus appropriée.

Une technologie de stockage multiniveau associée à la déduplication peut aider les entreprises à tirer parti d'une solution de stockage de données au meilleur coût, grâce à la mise en œuvre de la bonne technologie, au bon moment. Pour aller plus loin, de nombreuses entreprises, en réponse à la complexité des datacenters, se tournent à présent vers un modèle proactif de gestion des données basé sur le stockage multiniveau, qui inclut des niveaux de sauvegarde et des niveaux d'archivage actif couvrant l'intégralité des transferts de données intelligents de leurs workflows spécifiques.

Cinq caractéristiques essentielles à la déduplication

quelles sont les cinq caractéristiques que les solutions de déduplication modernes doivent proposer.

1. **Appliances de sauvegarde dédiées** : les datacenters de nouvelle génération sont d'une telle complexité qu'ils nécessitent des appliances de déduplication dédiées. Conçues pour fonctionner de pair avec un vaste éventail d'applications de sauvegarde, elles sont généralement simples à installer, offrent des performances optimales et jouent le rôle de disque cible pour les serveurs de sauvegarde. D'ici à 2018, Gartner prévoit que 50 % des applications soumises à des taux de modification élevés seront directement sauvegardées vers des appliances cibles de déduplication sans lien avec le serveur de sauvegarde, un chiffre qui atteint 10 % aujourd'hui*.
2. **Déduplication à partir de blocs de taille variable ou fixe** : la déduplication n'est pas une mince affaire. Les solutions de sauvegarde logicielles intégrant une fonction de déduplication s'appuient habituellement sur une approche à blocs fixes, car elle nécessite moins de calculs. Elle n'offre cependant généralement pas un niveau optimal de réduction des données. La déduplication à partir de blocs de taille variable est gourmande en ressources, mais minimise le stockage sur disque au fur et à mesure de l'accroissement des données et fournit le niveau de réduction de données le plus efficace du marché, tout en offrant une densité de stockage sur disque maximale. Les entreprises qui peuvent préférer la déduplication à partir de blocs de taille variable à l'approche par blocs fixes sont celles qui sont confrontées à une croissance rapide de leurs données ou qui intègrent des sites distants à protéger ou des environnements virtualisés. La déduplication à partir de blocs de taille variable peut également permettre de réduire le trafic réseau, une caractéristique essentielle à la réplication et à la reprise après incident.
3. **Évolutivité** : l'évolutivité à la demande offre une capacité de stockage facile à installer

pour une gestion simple et prévisible. Elle offre la possibilité aux utilisateurs d'augmenter la capacité à l'aide d'une simple clé de licence, sans autre installation sur site à prévoir. Les appliances virtuelles et physiques ont aujourd'hui des capacités allant de 1 To à plus de 500 To grâce à une approche évolutive, pour une augmentation de la capacité sans interruption, ou presque, des activités. Cette approche permet donc aux utilisateurs d'éviter le surdimensionnement de leur stockage et l'achat de capacité inutile.

4. **Supervision et création de rapports** : la déduplication est complexe si vous ne disposez pas des outils de gestion adaptés. La supervision et la génération de rapport proactives des fonctions de déduplication ne font pas souvent l'objet d'une attention suffisante. Elles permettent pourtant des décisions plus avisées et accélèrent la résolution des problèmes. Idéalement, le responsable d'un datacenter devrait être en mesure de superviser les sauvegardes depuis son appareil mobile. Les fonctionnalités avancées de création de rapports permettent à l'équipe informatique de gagner du temps pour se concentrer sur des projets plus stratégiques que la simple gestion des sauvegardes.
5. **Sécurité** : les entreprises étudient de plus en plus la sécurité de leurs données à tous les niveaux du workflow, dans le but d'éliminer les vulnérabilités avant qu'elles ne soient exploitées. Le chiffrement matériel AES 256 bits des données au repos et en transit offre une grande sécurité sans pour autant nuire aux performances. Il ne faut cependant pas oublier que les approches de chiffrement logiciel impliquent souvent une dégradation des performances.

Prêt pour la modernisation ?

Les tendances sont claires : à mesure que la taille et la complexité des datacenters augmentent, il est essentiel de passer de la simple déduplication à la *déduplication intelligente* qui prend en charge l'évolution des workflows de données.

*Gartner, *Magic Quadrant for Deduplication Backup Target Appliances*, 31 juillet 2014

crédit Photo @bluepix-shutterstock