

BUYER CASE STUDY

VMware améliore le taux d'utilisation du réseau et les performances de sauvegarde avec la déduplication EMC Avamar

Laura DuBois

ÉVALUATION D'IDC

Dans les environnements de datacenter soumis à des limites physiques, la technologie de déduplication permet de réaliser des économies significatives en optimisant le stockage et en réduisant l'encombrement du datacenter. Toutefois, une approche de la déduplication orientée client peut fournir d'autres avantages, en améliorant l'utilisation de la bande passante réseau et en minimisant les fenêtres de sauvegarde. Pour les entreprises qui utilisent la virtualisation de serveurs ou protègent les données des sites distants, l'approche orientée client permet en outre de diminuer les E/S réseau, de supprimer les sauvegardes des bureaux distants, d'éliminer les systèmes de sauvegarde sur bandes et de décharger les liaisons WAN encombrées. Cette étude de cas IDC :

- examine les avantages de la technologie de déduplication orientée client chez VMware Inc ;
- présente la mise en œuvre d'EMC Avamar au sein du datacenter de VMware ;
- aborde les prochains projets d'utilisation de la technologie dans les bureaux distants et les succursales de VMware.

DANS CETTE ÉTUDE DE CAS

Cette étude de cas IDC, qui s'appuie sur un entretien accordé en mars 2009 par Chanh Chi, architecte informatique chez VMware, et Michael Pate, directeur des opérations techniques chez VMware, examine l'utilisation faite par l'entreprise de la déduplication orientée client. En 2007, VMware a évalué et déployé les offres Avamar de déduplication et de réplication orientées client en vue de protéger et de restaurer les 300 systèmes physiques et virtuels de son infrastructure.

Ce document étudie la mise en œuvre réalisée par VMware de la technologie Avamar, y compris les difficultés rencontrées et ses besoins en termes d'évaluation. Il présente en outre les résultats de VMware et récapitule les futurs projets d'implémentation. Les utilisateurs peuvent s'appuyer sur ces informations pour déterminer si la technologie de déduplication orientée client est applicable à leur environnement et pour évaluer les avantages offerts par la solution Avamar.

PRÉSENTATION DE LA SITUATION

Présentation de l'entreprise

VMware (NYSE : VMW) est un des leaders mondiaux des solutions de virtualisation du poste de travail au datacenter. Les entreprises font confiance à VMware pour réduire leurs coûts, garantir la continuité d'activité, renforcer leur sécurité et adopter un fonctionnement plus écologique. Avec un chiffre d'affaires de 1,9 milliard de dollars en 2008, plus de 130 000 clients et plus de 22 000 partenaires, VMware figure parmi les éditeurs de logiciels cotés en bourse dont la croissance est la plus exceptionnelle. Il est détenu majoritairement par EMC Corp.

L'environnement informatique de VMware inclut des centaines de serveurs Windows et Linux physiques ou virtuels exécutant des applications telles que Microsoft Exchange, Microsoft SQL et Oracle, ainsi que des serveurs d'annuaires, de fichiers et d'imprimantes. L'entreprise gère environ 50 To de données sur l'ensemble de ses serveurs physiques et virtuels. Le département informatique prend en charge plus de 6 000 utilisateurs.

En tant que membres du département informatique de VMware, Chanh Chi et Michael Pate ont donné un aperçu des difficultés rencontrées par VMware, de l'évaluation de la nouvelle technologie de sauvegarde et des résultats de sa mise en œuvre à l'aide de la solution Avamar. En tant qu'architecte informatique et ancien administrateur de sauvegarde, M. Chi a activement participé à la définition des nouvelles directives et stratégies architecturales qui ont permis à l'entreprise de relever ses défis en termes de sauvegarde et de restauration.

Difficultés et solution

Le département informatique de VMware était confronté à plusieurs difficultés. Avec la croissance de son infrastructure, son datacenter fonctionnait dans un espace physique limité. L'environnement existant de l'entreprise utilisait EMC NetWorker pour les sauvegardes sur des bandes physiques et des appliances de déduplication Data Domain CIFS. Cette solution avait été mise en œuvre en vue de supprimer l'utilisation (et le retrait hors site) des 200 bandes physiques hebdomadaires. L'entreprise effectuait également une réplication vers le datacenter de colocation Savvis de Palo Alto en Californie, pour son site de reprise. Elle avait clairement besoin d'un datacenter plus important, plus stable et plus écologique. Elle avait également besoin d'une solution répondant à ses impératifs de protection des données actuels et futurs. Elle a donc décidé de réévaluer sa stratégie de sauvegarde actuelle et de redéfinir l'architecture de son environnement.

En examinant son environnement de sauvegarde, le département informatique de VMware a découvert de graves goulots d'étranglement dans son réseau. Comme on pouvait s'y attendre, l'entreprise utilisait ses solutions VMware pour consolider et optimiser le taux d'utilisation des serveurs disponibles. Plus de 95 % des systèmes VMware étaient virtualisés et les machines virtuelles invitées partageaient la même interface réseau physique sous-jacente sur les serveurs physiques. Les méthodes de sauvegarde existantes de VMware consommaient toute la capacité réseau disponible au cours de la fenêtre de sauvegarde 24 h/24 et 7 j/7, générant de graves goulots d'étranglement de sauvegarde. Avec un trop grand nombre de données à protéger et des fenêtres de sauvegarde de plus en plus courtes,

l'entreprise devait réaliser des sauvegardes complètes très volumineuses pour chaque client toutes les semaines, puis exécuter, au mieux, des sauvegardes incrémentielles le reste de la semaine. « Avec la croissance attendue des volumes de données, nous savions que cette approche aurait ses limites », explique M. Chi.

En 2006 (avant le rachat par EMC), VMware envisageait trois solutions pour résoudre ses problèmes de sauvegarde. L'entreprise avait d'abord pensé à intégrer dans son environnement de sauvegarde une nouvelle application de déduplication pour son environnement virtualisé. Elle envisageait également d'utiliser des solutions EMC Disk Library dans son environnement de sauvegarde. Toutefois, dans la mesure où ses principaux problèmes étaient la fenêtre de sauvegarde de plus en plus courte et le goulot d'étranglement réseau de son environnement virtualisé, l'entreprise étudiait principalement les solutions de déduplication orientées client, notamment les offres Symantec NetBackup PureDisk et EMC Avamar. À l'époque, la solution NetBackup PureDisk de Symantec ne prenait pas en charge tous les clients gérés dans l'environnement de VMware, notamment les serveurs Oracle et les filtres NAS, ce qui était rédhibitoire pour VMware.

L'équipe informatique de VMware réalisa alors une validation fonctionnelle sur site pour comparer les performances des trois scénarios de mise à niveau : sauvegardes EMC NetWorker sur bandes, sauvegardes EMC NetWorker sur appliances de déduplication Data Domain et sauvegardes EMC Avamar sur disques génériques. L'équipe demanda une unité d'évaluation Avamar et la testa dans son propre environnement pendant environ deux mois. Elle réalisa plusieurs tests pour comparer les résultats des sauvegardes des 12 mêmes serveurs de fichiers Windows et Linux et ceux d'une sauvegarde Exchange, soit un volume total de données d'environ un demi-téraoctet. Les tests démontrèrent clairement que la sauvegarde Avamar était plus performante que les autres solutions malgré sa connexion réseau plus lente. « Les taux de déduplication étaient variables, mais en moyenne, nous avons constaté un taux d'optimisation 3 fois plus élevé lors de la première sauvegarde de la plupart des clients. Pour certains systèmes de fichiers Windows, ce taux était même 500 fois supérieur pendant les sauvegardes suivantes », rapporte M. Chi.

Résultats

Au vu des tests de validation fonctionnelle très probants, VMware a choisi le produit Avamar et l'a déployé en décembre 2007. Bien que VMware fasse déjà partie d'EMC à l'époque, le déploiement d'Avamar fut réalisé après le PAPE de VMware, avec pour conséquence pour l'entreprise l'obligation d'acheter le produit. Le nouveau datacenter de production installé à Santa Clara, en Californie, est entré en activité fin 2007. « La fonction de réplication à distance d'Avamar a joué pour beaucoup dans notre décision. Elle nous a permis de répliquer les données de sauvegarde entre notre nouveau datacenter de production et le datacenter de reprise après sinistre, supprimant ainsi la gestion des bandes », témoigne M. Chi.

VMware a acquis deux systèmes Avamar composés de 12 nœuds de données chacun. Les experts EMC Professional Services ont aidé VMware à configurer Avamar pendant deux jours. Un système Avamar est installé à Palo Alto, en Californie, et les données sont répliquées quotidiennement et de manière bidirectionnelle entre les deux systèmes Avamar. VMware a commencé avec 12 nœuds de données dans chaque « grille » Avamar, la première grille protégeant environ 12 To de données principales et la seconde environ 7 To. Sur

les deux grilles, la solution Avamar protège plus de 300 serveurs de fichiers physiques et virtuels. L'entreprise a conservé EMC NetWorker pour la sauvegarde directe des serveurs de fichiers EMC Celerra sur bandes et pour la sauvegarde de l'environnement Exchange sur des périphériques Data Domain existants. Étant donné que de plus en plus de clients se connectent, le département informatique de VMware a depuis ajouté quatre nouveaux nœuds de stockage par grille, soit une capacité de 16 To par grille, et a acquis une grille de densité supérieure, d'une capacité de 32 To :

- ☒ **Sauvegardes des machines virtuelles.** VMware déploie les agents logiciels Avamar sur chaque machine virtuelle invitée et déploie Avamar dans son environnement VMware Consolidated Backup (VCB) pour quelques clients qui requièrent une protection supplémentaire. Dans tous les cas, Avamar permet des sauvegardes quotidiennes complètes et rapides.
- ☒ **Fenêtre de sauvegarde.** « Au bout de 10 minutes, la moitié des 360 clients ont exécuté leur sauvegarde complète quotidienne. Au bout d'une heure, seuls quelques-uns ne sont pas encore sauvegardés », indique M. Chi. Avant la mise en œuvre d'Avamar, les sauvegardes pouvaient durer une journée, voire plus.
- ☒ **Performances de sauvegarde.** Avec un seul client virtuel et 12 LUN de 415 Go de données chacune, la sauvegarde avec NetWorker prendrait une semaine. Désormais, une sauvegarde Avamar de 5 To et 40 à 50 millions de fichiers prend environ 11 heures.
- ☒ **Performances réseau.** De manière générale, l'entreprise sauvegarde environ 18 To de données la nuit et gère un trafic réseau de 200 Go, utilisant environ 1 % de la bande passante réseau nécessaire pour une sauvegarde traditionnelle.
- ☒ **Performances de restauration.** Utiliser Avamar pour restaurer les données est considérablement plus simple et plus rapide qu'avec les solutions précédentes. « Auparavant, la restauration de fichiers après une corruption de données ou une erreur d'un utilisateur prenait quelques heures pour un fichier de 200 Ko. Avec Avamar, elle dure quelques secondes. Les données sont « réhydratées » sur le client et il est fréquent de voir un fichier de 8 Go restauré en seulement 30 secondes », explique M. Chi.
- ☒ **Interface utilisateur.** « L'interface utilisateur d'Avamar est très intuitive et aucune véritable expérience de la sauvegarde n'est nécessaire. Les tâches de sauvegarde peuvent ainsi être réparties entre un plus grand nombre de personnes. Par exemple, les équipes d'ingénieurs peuvent réaliser leur propre sauvegarde et restauration du code source », ajoute M. Chi.

VMware conserve actuellement les sauvegardes pendant environ 30 jours sur les systèmes à disques Avamar et vide toutes les semaines les sauvegardes sur des bandes pour la rétention à long terme. L'entreprise effectue une troisième copie en dernier ressort.

Projets futurs

« Nous avons doublé notre capacité Avamar au cours de la semaine passée. Nous avons acquis un système Avamar d'une capacité de 52 To pour

sauvegarder nos sites R&D et techniques distants dans le monde entier. Nous avons intégré une grille dans un nouveau datacenter situé dans l'état de Washington et capable de sauvegarder 10 sites et succursales distants », relate à son tour M. Pate. VMware utilisera l'agent logiciel Avamar local pour les sauvegardes sur un datacenter plus important. Dans ce cas, la solution Avamar a facilité le passage de liaisons WAN limitées au datacenter régional plus important. Cette approche a permis à VMware de sauvegarder les principaux serveurs de fichiers des sites distants vers une grille Avamar Data Store au sein d'un datacenter plus important. Elle a également permis de supprimer l'utilisation des changeurs de bandes sur les sites distants, éliminant par là même le coût lié à environ 1 000 bandes par trimestre.

CONSEILS ESSENTIELS

Conseils aux utilisateurs

Les utilisateurs devraient suivre les recommandations suivantes pour évaluer les avantages potentiels de la technologie de déduplication des données au sein de leur environnement :

- ☒ **En matière de déduplication, l'égalité n'existe pas.** La technologie de déduplication peut être utilisée en différents points, tels que les logiciels de sauvegarde, les dispositifs de sauvegarde, etc. Les utilisateurs doivent déterminer les principales difficultés auxquelles ils sont confrontés et, ainsi, identifier l'approche de déduplication la plus adaptée à leurs besoins.
- ☒ **Nécessité d'une validation fonctionnelle.** La déduplication est très prisée actuellement, et de nombreux fournisseurs prétendent proposer une technologie de déduplication. Or, IDC a mis en évidence, au cours de validations fonctionnelles, que certaines solutions de déduplication ne fonctionnent pas comme prévu.
- ☒ **Test de la sauvegarde des données en conditions réelles.** Les taux de déduplication varient en fonction du type de données à sauvegarder. Certaines données se dédupliquent facilement et d'autres non. Connaître les taux de déduplication réels tels que définis dans le présent document, ainsi que les taux de déduplication obtenus lors de votre validation fonctionnelle, vous évitera les mauvaises surprises une fois la technologie mise en œuvre.
- ☒ **Considérer d'autres formes d'optimisation de la charge de travail.** Les technologies d'optimisation, telles que la compression et la déduplication, réduisent les coûts liés à l'utilisation de disques pour la protection des données et permettent de réaliser un plus grand nombre de sauvegardes sur disques sur de longues périodes. Les entreprises doivent s'assurer que la compression et la déduplication (ainsi que le cryptage) sont utilisées de manière complémentaire et dans le bon ordre pour pouvoir profiter des avantages offerts par la technologie d'optimisation.

EN SAVOIR PLUS

Études associées

- ☒ *Worldwide Storage Software 2009–2013 Forecast: Quest for Efficiency Driving Investments in Storage Software* (IDC #217529, mars 2009)

- ☒ *Worldwide Data Protection and Recovery Software 2007 Vendor Shares: Market Rebounds* (IDC #213320, juillet 2008)

Copyright

Cette étude IDC a été publiée dans le cadre d'un service de veille stratégique d'IDC qui propose des recherches écrites, des interactions entre analystes, des télébriefings et des conférences. Visitez le site www.idc.com pour en savoir plus sur les services d'abonnement et de conseils d'IDC. Pour afficher la liste des bureaux d'IDC dans le monde, consultez le site www.idc.com/offices. Contactez la hotline d'IDC au 800.343.4952, numéro de poste : 7988 (ou +1.508.988.7988) ou envoyez un courriel à sales@idc.com pour plus d'informations sur le prix de l'abonnement à un service IDC, sur l'envoi de copies supplémentaires ou sur les droits Web.

Copyright 2009 IDC. Toute reproduction interdite sauf autorisation. Tous droits réservés.