

Avec les Nuls, tout devient facile !

Une édition spéciale de HPE SimpliVity

L'infrastructure hyperconvergée

**POUR
LES NULS[®]**

Apprenez à :

- Utiliser une infrastructure hyperconvergée pour simplifier l'informatique et réduire le TCO
- Utiliser une infrastructure hyperconvergée pour simplifier l'informatique et réduire le coût total de possession
- Reproduire le modèle économique et l'agilité des très grandes entreprises Web dans votre datacenter

Proposé par



**Hewlett Packard
Enterprise**

Scott D. Lowe

Consultant et vétéran de l'industrie



À propos de Hewlett Packard Enterprise

Hewlett Packard Enterprise est une entreprise leader dans l'industrie des technologies qui permet à ses clients d'aller plus loin et plus vite. Avec l'offre la plus complète du marché, allant du cloud au datacenter en passant par les applications professionnelles, nos technologies et services aident nos clients à travers le monde à améliorer l'efficacité, la productivité et la sécurité de leurs systèmes d'information.



**Hewlett Packard
Enterprise**

L'infrastructure hyperconvergée

POUR

LES NULS[®]

Une édition spéciale de HPE SimpliVity

par Scott D. Lowe

WILEY

L'infrastructure hyperconvergée pour les Nuls®, une édition spéciale de HPE SimpliVity

Édité par
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2018 par John Wiley & Sons, Inc., Hoboken, New Jersey

Aucune partie de cet ouvrage ne peut être reproduite, conservée dans un système d'extraction, ou transmise sous quelque forme ou par quelque moyen que ce soit, par voie électronique ou mécanique, photocopie, enregistrement, numérisation ou autre, sans l'accord écrit préalable de l'éditeur, sauf si les articles 107 et 108 de la loi des États-Unis de 1976 relative au droit d'auteur (« United States Copyright Act ») l'autorisent. Les demandes d'autorisation auprès de l'éditeur doivent être adressées à Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, ou en ligne à l'adresse <http://www.wiley.com/go/permissions>.

Marques commerciales :Wiley, Pour les Nuls, le logo Dummies Man, The Dummies Way, Dummies.com, Avec les Nuls, tout devient facile !, et les appellations commerciales afférentes sont des marques de commerce ou des marques déposées de John Wiley & Sons, Inc. et/ou de ses sociétés affiliées aux États-Unis et dans d'autres pays, et ne peuvent pas être utilisées sans autorisation écrite. SimpliVity et le logo de SimpliVity sont des marques de commerce de la SimpliVity Corporation. Toutes les autres marques déposées sont la propriété de leurs propriétaires respectifs. John Wiley & Sons, Inc. n'est associé à aucun produit ou distributeur mentionné dans cet ouvrage.

LIMITATION DE RESPONSABILITÉ/EXCLUSION DE GARANTIE : L'ÉDITEUR ET L'AUTEUR NE FONT AUCUNE DÉCLARATION NI N'ACCORDENT AUCUNE GARANTIE QUANT À L'EXACTITUDE OU À L'EXHAUSTIVITÉ DU CONTENU DU PRÉSENT OUVRAGE. ILS REJETTENT EN PARTICULIER TOUTES LES GARANTIES, Y COMPRIS, SANS AUCUNE LIMITE, LES GARANTIES D'ADÉQUATION À UN USAGE PARTICULIER. AUCUNE GARANTIE NE PEUT ÊTRE CRÉÉE OU PROROGÉE PAR DES DOCUMENTS DE VENTE OU DE PROMOTION. LES CONSEILS ET STRATÉGIES CONTENUS DANS LE PRÉSENT OUVRAGE PEUVENT NE PAS CONVENIR À TOUTES LES SITUATIONS. LE PRÉSENT OUVRAGE EST VENDU ÉTANT ENTENDU QUE L'ÉDITEUR N'OFFRE PAS DE SERVICES JURIDIQUES, COMPTABLES OU AUTRES SERVICES PROFESSIONNELS. LES LECTEURS QUI VEULENT OBTENIR UNE ASSISTANCE PROFESSIONNELLE DOIVENT S'ADRESSER À UN PROFESSIONNEL COMPÉTENT. NI L'ÉDITEUR NI L'AUTEUR NE SERA TENU RESPONSABLE DES DOMMAGES DÉROULANTS DU CONTENU DU PRÉSENT OUVRAGE. LA MENTION D'UNE ORGANISATION OU D'UN SITE INTERNET DANS LE PRÉSENT OUVRAGE, EN CITATION ET/OU COMME SOURCE POTENTIELLE DE RENSEIGNEMENTS COMPLÉMENTAIRES NE SIGNIFIE PAS QUE L'AUTEUR OU L'ÉDITEUR ENTÈRE LES INFORMATIONS OU LES RECOMMANDATIONS QUE PEUT FOURNIR L'ORGANISATION OU LE SITE INTERNET. EN OUTRE, LES LECTEURS DOIVENT SAVOIR QUE LES SITES INTERNET MENTIONNÉS DANS LE PRÉSENT OUVRAGE PEUVENT AVOIR CHANGE OU DISPARU DEPUIS LA DATE DE RÉDACTION DE CET OUVRAGE.

Pour obtenir des renseignements généraux sur nos autres produits et services, ou sur la publication d'un livre sur mesure *Pour les Nuls* destiné à votre entreprise ou organisation, veuillez contacter notre service de développement commercial aux États-Unis, par téléphone au 877-409-4177, par e-mail à info@dummies.biz, ou consulter notre site www.wiley.com/go/custompub. Pour obtenir des informations sur les licences relatives à la marque *Pour les Nuls* pour des produits et services, veuillez écrire à l'adresse BrandedRights&Licenses@wiley.com.

ISBN 978-1-119-47864-5 (pbk) ; ISBN 978-1-119-47865-2 (ebk)

Imprimé aux États-Unis d'Amérique

10 9 8 7 6 5 4 3 2 1

Remerciements de l'éditeur

Cet ouvrage a été réalisé avec la participation des personnes suivantes :

Éditeur de développement : Kathy Simpson

Rédactrice projet : Zoë Wykes

Rédacteur chargé des acquisitions :
Katie Mohr

Responsable éditorial : Rev Mengle

Représentante du développement commercial : Sue Blessing

Coordnatrice du projet : Melissa Cossell

Contributeurs de HPE SimpliVity :
Jesse St. Laurent, Lauren Whitehouse

Table des matières

Introduction	1
À propos de cet ouvrage.....	1
Quelques suppositions	2
Icônes employées dans cet ouvrage	2
Au-delà de ce livre	2
Et maintenant ?	2
Chapitre 1 : Les bases de l'hyperconvergence	3
Définition de l'infrastructure hyperconvergée.....	3
Systèmes d'hyperconvergence	4
Chapitre 2 : Les problèmes de la virtualisation	5
Infrastructure innovatrice	5
Ressources sous-utilisées.....	6
Interfaces de gestion multiples	7
Difficultés et retards de déploiement.....	8
Stockage	10
L'effet « I/O blender ».....	10
Boot storm de la VDI.....	12
Atténuation	12
Points de contact multiples	13
Décalage avec les politiques	16
Chapitre 3 : Bienvenue dans le software-defined datacenter (SDDC)	17
Virtualisation.....	17
Automatisation	18
IT as a Service	19
Chapitre 4 : Ce que les entreprises attendent de l'informatique	21
Plus d'efficacité.....	22
Meilleure utilisation du temps	22
Mise en correspondance des compétences et des tâches	22
Bonne gestion des ressources	23
Respect des budgets.....	24
Diminution des risques	24
Plus d'agilité.....	25

Chapitre 5 : Comment le cloud fait évoluer l'informatique	27
Évolutivité et partage des ressources.....	27
Conception axée sur les logiciels.....	28
Économies d'échelle.....	28
Flexibilité des ressources.....	29
Rendre l'automatisation possible.....	29
Retirer la politique de l'infrastructure	29
Adoption d'une approche centrée sur les VM.....	30
Comprendre l'économie du cloud.....	30
Chapitre 6 : Compréhension de l'infrastructure hyperconvergée	33
L'évolution de la convergence.....	34
Systèmes intégrés.....	34
L'infrastructure convergée	34
L'infrastructure hyperconvergée.....	35
Caractéristiques de la convergence.....	36
Chapitre 7 : Les neuf choses qu'une infrastructure hyperconvergée peut faire pour vous.	37
Focalisation sur les logiciels	38
Systèmes et gestion centralisés	38
Plus d'agilité.....	39
Évolutivité et efficacité.....	39
Une infrastructure rentable.....	40
Automatisation facile.....	40
Focalisation sur les VM	40
Ressources partagées	41
Protection des données	42
Chapitre 8 : Sept façons d'appliquer l'hyperconvergence	43

qui découlent de la mise en œuvre d'un datacenter basée sur une infrastructure hyperconvergée.

Quelques suppositions

Dans ce livre, je présume que vous avez quelques notions des tendances et des modèles de virtualisation et de cloud computing. Cet ouvrage s'adresse donc avant tout aux dirigeants et cadres de services informatiques (directeurs des systèmes d'information, directeurs de la technologie, directeurs de service informatique et responsables techniques).

Icônes employées dans cet ouvrage

Vous trouverez plusieurs icônes dans les marges de ce livre. Voici ce qu'elles veulent dire.



Il est utile de mémoriser les éléments signalés par l'icône « Rappel ».



L'icône « Jargon technique » signale une lecture plus poussée. Vous pouvez sauter cette section si vous le souhaitez (mais j'espère que non).



L'icône « Conseil » signale des informations utiles.



L'icône « Attention » vous alerte à l'égard de risques divers.

Au-delà de ce livre

Ce sujet est bien trop vaste pour le traiter en détail en si peu de pages. Si vous avez envie d'en savoir plus sur l'infrastructure hyperconvergée après la lecture de cet ouvrage, consultez le site www.hpe.com/info/hc.

Et maintenant ?

Comme tous les livres *Pour les Nuls*, cet ouvrage peut être lu dans l'ordre de votre choix. Commencez par le chapitre qui vous intéresse le plus ou lisez-le de bout en bout. Après tout, c'est votre livre, c'est à vous de voir.

Chapitre 1

Les bases de l'hyperconvergence

Dans ce chapitre

- ▶ Comprendre en quoi consiste une infrastructure hyperconvergée
- ▶ Aborder les nombreuses formes de l'hyperconvergence

De temps à autre, la technologie d'entreprise subit une évolution énorme lors de l'apparition de nouveaux modèles qui répondent à l'évolution des besoins des entreprises. Ce chapitre porte sur l'infrastructure hyperconvergée qui est l'aboutissement et la réunion de plusieurs tendances qui apportent une valeur particulière à l'entreprise moderne.

Définition de l'infrastructure hyperconvergée

L'hyperconvergence, de quoi s'agit-il ? Au plus haut niveau, c'est un moyen de permettre une économie à l'échelle du cloud sans compromettre les performances, la résilience et la disponibilité que vous attendez de votre datacenter. L'infrastructure hyperconvergée offre d'importants avantages :

- ✓ **Efficacité des données** : l'infrastructure hyperconvergée réduit les besoins de stockage, de largeur de bande et d'IOPS.
- ✓ **Élasticité** : l'infrastructure hyperconvergée facilite l'évolution/réduction de ressources en fonction des demandes de l'entreprise.
- ✓ **Centrage sur la charge de travail** : l'attention est portée sur la charge de travail (ou machine virtuelle) en tant que pierre angulaire des services de l'information de l'entreprise, avec tous les systèmes associés orientés sur des charges de travail uniques.
- ✓ **Protection des données** : assurer la restauration de données en cas de perte ou de corruption est une exigence fondamentale en informatique, que l'infrastructure hyperconvergée rend plus facile.
- ✓ **Mobilité entre machines virtuelles** : l'infrastructure hyperconvergée facilite encore davantage la mobilité entre les applications/charges de travail.

- ✓ **Résilience** : l'infrastructure hyperconvergée permet des niveaux plus élevés de disponibilité des données qu'avec les systèmes existants.
- ✓ **Rentabilité** : l'infrastructure hyperconvergée offre un modèle économique durable par étapes qui élimine le gaspillage en informatique.

Systèmes d'hyperconvergence

La convergence prend des formes multiples. À son niveau le plus basique, la convergence se limite à réunir les commutateurs individuels de stockage, calcul et réseau en solutions prétestées et prévalidées qui sont vendues comme une seule solution. Mais ce niveau de convergence ne fait que simplifier le cycle d'achat et de mise à niveau. Il ne permet pas de relever les défis opérationnels constants qui accompagnent souvent le passage à la virtualisation. Il faut quand même créer des LUN, acquérir et configurer des optimiseurs de réseau WAN, puis acheter et entretenir des produits indépendants de sauvegarde et de reproduction.

L'infrastructure hyperconvergée combine harmonieusement les services de calcul, stockage, mise en réseau et gestion de données en une seule solution, un seul système physique. Le logiciel qui permet l'hyperconvergence s'exécute sur des systèmes x86 standards, dans le but d'exécuter des charges de travail virtualisées ou conteneurisées. L'architecture distribuée vous permet de réunir plusieurs systèmes sur un même site ou entre plusieurs, et forme un pool de ressources partagées qui facilite une grande disponibilité, la mobilité des charges de travail et l'évolutivité efficace des performances et des capacités. Gérées normalement par l'intermédiaire d'une seule interface, les infrastructures hyperconvergées vous laissent définir votre politique et exécuter les activités au niveau de la machine virtuelle/du conteneur.

Les résultats sont de taille : entre autres, la baisse des dépenses en capital en conséquence de coûts initiaux d'infrastructure réduits, la baisse des frais d'exploitation grâce à la réduction des coûts opérationnels et de personnel et un seuil de rentabilité lié aux nouveaux besoins de l'entreprise atteint plus rapidement. D'un point de vue technique, les tout nouveaux généralistes de l'informatique (les informaticiens qui ont des connaissances larges des besoins d'infrastructure et de l'entreprise) n'ont aucune difficulté à prendre en charge des systèmes hyperconvergés. Les organisations n'ont plus besoin de disposer d'ingénieurs spécialisés pour gérer chaque aspect du datacenter.

Si l'on veut parfaitement comprendre l'hyperconvergence, il est important d'appréhender les tendances qui ont mené l'industrie jusqu'ici. Elles comprennent les casse-têtes de la post-virtualisation, la montée du software-defined datacenter et le cloud computing.

Chapitre 2

Les problèmes de la virtualisation

Dans ce chapitre

- ▶ Évolution au rythme des innovations
- ▶ Utilisation rationnelle des ressources
- ▶ Examen des interfaces de gestion
- ▶ Discussion sur les points de contact
- ▶ Établissement de politiques efficaces

La virtualisation a changé l'informatique et le datacenter d'une manière fondamentale et permanente, c'est un fait. Aujourd'hui, la plupart des services s'exécutent au sein d'environnements virtuels et l'informatique adopte souvent une approche « virtualisée avant tout » face au déploiement de nouvelles applications et nouveaux services. Cela veut dire que les administrateurs envisagent un environnement virtuel pour l'exécution de nouvelles applications plutôt que de construire un environnement physique neuf.

Bien que la virtualisation offre d'importants avantages, elle s'accompagne de défis que l'informatique doit relever pour que l'entreprise puisse aller de l'avant. Ce chapitre décrit ces défis.

Infrastructure innovatrice

À chaque fois qu'une start-up met un nouveau produit de grande qualité sur le marché, les entreprises se démènent pour mettre en œuvre cette solution. La prolifération de périphériques conçus spécifiquement pour un usage a engendré un niveau de complexité inutile dont le résultat est le chaos au sein du datacenter.

L'innovation c'est merveilleux, et nous voulons tous qu'elle perdure, mais à force, les datacenters sont tellement surchargés qu'ils deviennent ingérables. Il est temps de faire du vide, pour ainsi dire.

Ces dix dernières années, les services informatiques ont porté leur attention sur le problème des capacités de stockage et ils ont déployé toutes sortes de technologies pour dompter le fauve de la capacité, comme l'optimisation du réseau WAN et les dispositifs de déduplication de sauvegarde. Par conséquent, les technologies de gestion efficace des données sont devenues la norme dans beaucoup de produits.

Mais que se passe-t-il si l'on met ces produits ensemble dans le datacenter ? On ne cesse de dédupliquer et hydrater des données alors qu'elles passent d'un périphérique à un autre. Le stockage engendre la déduplication des données ; puis on lit les données à sauvegarder qui doivent alors être hydratées (pour les remettre dans un état que l'application de sauvegarde comprend) et souvent dédupliquées une nouvelle fois quelque part sur le chemin de sauvegarde des données. Pour le processeur, le coût associé au retraitement des mêmes données est énorme, sans oublier le coût en largeur de bande de toutes les données hydratées.



La déduplication est le processus d'examen des données pour détecter des blocs communs. Une fois qu'ils sont identifiés, les blocs communs sont remplacés par un pointeur minuscule renvoyant à une copie unique des données déjà stockées sur le disque — ce qui demande nettement moins de capacité si elles sont enregistrées sur le support de stockage. La déduplication permet d'énormes économies de capacité de stockage et, tout aussi important, au niveau des opérations d'entrée-sortie par seconde (IOPS) car moins d'écritures sont faites sur le disque. *L'hydratation* inverse le processus de déduplication, comme quand on transfère des données sur un nouveau système qui ne supporte pas les données dédupliquées.

Je parle de ce problème plus en détail plus tard dans ce chapitre.

Ressources sous-utilisées

La virtualisation a permis aux organisations de consolider bon nombre de leurs serveurs qui s'exécutent sur une plateforme commune : la couche logicielle de l'hyperviseur. Cette évolution a permis aux services informatiques de faire bien meilleur usage des serveurs à leur disposition. Avant la virtualisation, il était fréquent qu'un serveur atteigne une moyenne d'utilisation de juste 15 %. La virtualisation a fait beaucoup grimper cette moyenne. Résultat : les organisations bénéficient maintenant d'un retour nettement supérieur sur leurs investissements en serveurs. De plus, elles n'ont généralement pas besoin d'acheter autant de serveur physiques que par le passé.

La virtualisation a changé la donne au niveau des besoins en serveurs. Malheureusement, les services informatiques doivent souvent conserver des groupes différents de personnes pour gérer des ressources matérielles distinctes. Un groupe s'occupe du stockage par exemple, un autre gère l'aspect serveurs et un troisième la mise en réseau. Si un problème survient, il arrive souvent qu'on se renvoie la balle.

Par ailleurs, les charges de travail émergentes posent des problèmes de ressources qui poussent les services informatiques à développer des environnements d'infrastructure par service. Les environnements VDI (infrastructure de bureau virtuel) par exemple, produisent des schémas d'utilisation des ressources très différents de ceux des projets de virtualisation de serveurs. Pour répondre aux attentes des utilisateurs en matière de VDI, les professionnels de l'informatique déploient souvent des environnements complètement séparés, des serveurs jusqu'au stockage.

Les problèmes d'îlots de ressources ne sont-ils pas exactement ce que la virtualisation est censée éliminer ? Ces îlots sont en grande partie responsables de la sous-utilisation. La virtualisation devrait engendrer un pool unique de ressources spécifiquement conçues pour répondre aux besoins des applications et de fait, en maximiser l'utilisation.

Interfaces de gestion multiples

Supports de stockage. Optimisateurs. Hyperviseurs. Équilibrateurs de charge. Qu'ont-ils en commun ? Chacun de ces composants distincts a sa propre interface de gestion. Si vous utilisez plusieurs composants, chacun accompagné d'une console distincte de gestion (et d'un moteur de politiques) au lieu d'un système administratif unique, centralisé et convivial, vous serez confronté aux problèmes suivants :

- ✓ Des fournisseurs qui s'accusent réciproquement si quelque chose ne marche pas.
- ✓ L'incapacité à faire évoluer votre environnement de datacenter d'une manière facile et linéaire.
- ✓ Un niveau de complexité accru en raison des politiques et une gestion liée à des composants informatiques plutôt qu'à des charges de travail.

Difficultés et retards de déploiement

Si les organisations continuent à rencontrer des problèmes quand elles déploient de nouvelles applications et de nouveaux services, les enjeux liés aux ressources en sont la cause principale. Les coûts administratifs les suivent de près. Permettez-moi de vous expliquer pourquoi.



Passage à une structure horizontale

Le datacenter existant est très délicat de bien des manières. Un changement à un niveau quelconque peut perturber toute la structure. Les fournisseurs d'hyperconvergence ont tiré les enseignements et repris les tactiques des fournisseurs cloud et remplacé les datacenters composés de plusieurs niveaux ou de silos par une structure du système informatique beaucoup plus plate. Comme pratiquement tout le matériel autrefois séparé du datacenter est réuni dans l'environnement hyperconvergé, le service informatique doit porter son attention ailleurs, en modifiant ses structures de gestion de ressources et ses compétences. Plutôt que de disposer d'un personnel d'experts très compétents dans chaque domaine, l'hyperconvergence peut donner lieu à des généralistes de l'infrastructure.

Les connaissances des généralistes de l'infrastructure qui émergent sont générales sur l'ensemble des ressources plutôt qu'approfondies sur des ressources particulières. Une connaissance approfondie n'a aucune utilité. Dans un monde hyperconvergé, tout ce qui est complexe est traité sous le capot. Il ne faut que des connaissances générales aux généralistes

de l'infrastructure pour répondre aux besoins de l'entreprise et gérer tout l'environnement à travers une seule interface d'administration. De bien des façons, ces personnes se concentrent beaucoup plus sur les applications que leurs prédécesseurs axés sur des îlots. Il leur suffit de savoir comment appliquer des ressources d'infrastructure pour répondre aux besoins individuels d'applications.

Pour les services informatiques qui avaient du mal à faire correspondre les opérations IT aux besoins de l'entreprise, ce développement est synonyme de très bonnes nouvelles sur plusieurs points :

- Cette nouvelle structure ouvre la voie à l'élimination des îlots inefficaces de gestion de ressources qui ont vu le jour en informatique.
- Un datacenter d'une structure horizontale géré par un groupe de techniciens de l'infrastructure permet des économies d'échelle par rapport à la gestion antérieure des ressources en îlots.
- Les généralistes de l'infrastructure sont bien plus près des applications que ne l'étaient les spécialistes.



Les baies flash corrigent des problèmes spécifiques

Avec l'émergence du stockage flash (et vraiment rapide) à des prix relativement abordables, de nouvelles branches sont nées sur le marché du stockage. Une de ces branches offre des baies de stockage uniquement basées sur le stockage flash.

Bien que les fournisseurs de cet espace 100 % flash proposent des produits vraiment attrayants, ils sont pour beaucoup conçus pour résoudre les problèmes d'une seule application — pensez VDI et analyse de big data. Si l'on considère les charges de travail de nombreuses entreprises, les baies de stockage AFA (All-Flash Array) ne font, par définition, que répondre à un problème de performances par du matériel. Les solutions de stockage qui combinent le stockage flash et des

disques rotatifs peuvent constituer une approche plus équilibrée et raisonnable pour répondre aux besoins liés aux charges de travail. De plus, le coût par gigaoctet d'un stockage flash est plutôt onéreux par rapport à d'autres options. Ceci dit, pour les applications qui doivent réaliser des centaines de milliers, voire des millions d'IOPS dans un minuscule espace, les baies AFA sont imbattables. Pour tout le reste, envisagez des options de stockage plus équilibrées. N'oubliez pas, une baie de stockage flash offre des performances de stockage supérieures. Elle ne résout ni le problème des îlots de ressources, ni celui de la gestion de l'infrastructure, ni les défis d'interopérabilité, ni les difficultés d'évolutivité du datacenter moderne.

Concernant les ressources, les enjeux sont multiples, entre autres :

- **L'effet « I/O blender »** : la consolidation de machines virtuelles (VM) contribue à une charge de travail aléatoire I/O (entrées et sorties) — chacune avec un schéma individuel de lecture/écriture de données sur le support de stockage. Je parle de l'effet *I/O blender* en détail plus tard dans ce chapitre.
- **Capacité** : un autre enjeu consiste à s'assurer que la capacité évolue à mesure que l'organisation s'agrandit. Dans un contexte de ressources réparties et d'îlots de ressources éparpillés dans le datacenter, la gestion continue de la capacité se complique encore si l'on veut qu'elle soit suffisante.
- **Coût** : même si vous avez assez de ressources pour déployer une nouvelle application (voir le point qui précède), le coût administratif de ce processus engendre ses propres problèmes :
 - Un nouveau LUN (logical unit number) doit être établi pour supporter la nouvelle application. Ce processus

peut demander plusieurs étapes s'il y a plusieurs niveaux de stockage.

- Une ou plusieurs VM doivent être provisionnées.
- La mise en réseau de ces nouvelles VM doit être configurée.
- Des équilibrateurs de charges et des dispositifs d'optimisation du réseau WAN (réseau étendu) doivent être gérés pour supporter les nouvelles VM.
- Des mécanismes de protection des données doivent être mis en œuvre pour les nouveaux services.

Misère ! Ça en fait des choses à faire ! Toutes qui demandent du temps et qui impliquent des équipes de personnes différentes du service informatique. Je vous souhaite bonne chance !

Stockage

La virtualisation dépend lourdement du stockage, mais cette ressource a semé la zizanie dans les entreprises qui mettent tout en œuvre pour passer à la virtualisation à 100 %. Voici pourquoi.

Réfléchissez aux charges de travail traditionnelles basées sur des serveurs physiques. Au moment de bâtir ces environnements d'applications, chaque serveur était soigneusement adapté pour répondre aux configurations uniques de chaque application individuelle. Les serveurs de bases de données avaient deux disques — un pour les fichiers de la base de données et un pour les fichiers journaux — avec des structures différentes d'un ensemble redondant de disques indépendants (RAID). Les serveurs de fichiers avaient des structures RAID de niveau 5 pour optimiser la capacité tout en maintenant la protection des données.

Maintenant, prenez votre environnement virtualisé. Vous avez amassé tous ces environnements d'applications bâtis soigneusement dans un environnement unique à stockage partagé. Chaque application a des besoins de stockage particuliers, mais vous avez tout simplement demandé au stockage de tout organiser pour vous et il n'a pas toujours fait du bon boulot.

L'effet « I/O blender »

Les systèmes de stockage traditionnels étaient optimisés autour de la gestion des LUN. Les LUN étaient reproduits d'un contrôleur dans une baie de stockage à un LUN lié au contrôleur d'une autre baie. Les systèmes de stockage prenaient un cliché des LUN et ceux-ci pouvaient être déplacés d'un serveur à un autre.

De nos jours, les serveurs ont été remplacés par des VM et bon nombre d'entre elles fonctionnent sur un seul serveur et un grand nombre de serveurs utilisent un LUN unique pour stocker les VM. Cela signifie que le système de stockage a des dizaines (voire des centaines) de serveurs logiques (VM) tous stockés sur le même LUN. La gestion d'une application, d'un serveur ou d'une VM unique n'est plus possible du point de vue du système de stockage.

Une plateforme axée sur les VM élimine cet effet I/O blender (un terme qui sert à décrire les environnements dans lesquels des schémas combinés d'entrées et sorties sont en concurrence pour accéder à des ressources de stockage limitées) et vous permet d'optimiser des VM individuelles. Des politiques peuvent être appliquées à des VM individuelles. Les performances de VM individuelles peuvent être optimisées. Les sauvegardes des VM peuvent être gérées individuellement et la reproduction est configurée par VM.

Voyez-vous le schéma qui se dessine ici ?

Quand toutes vos applications essaient de fonctionner ensemble sur le même LUN, cela crée un effet I/O blender. Voici certaines façons dont des services courants contribuent à l'effet I/O blender :

- **Bases de données** : les bases de données sont caractérisées par des schémas I/O aléatoires. Le système doit sauter d'un point à un autre du disque pour trouver ce que vous cherchez.
Fichiers journaux de la base de données : les fichiers journaux sont séquentiels par nature. Normalement, un enregistrement sur les fichiers journaux suffit encore une fois, d'une manière séquentielle.
- **Stockage de fichiers aléatoire** : les serveurs de fichiers sont très aléatoires au niveau des entrées et des sorties. On ne sait jamais quand un utilisateur va enregistrer un nouveau fichier ou en ouvrir un ancien.
- **Applications au niveau de l'entreprise** : les applications comme Microsoft Exchange et SharePoint sont sensibles sur le plan de la configuration du stockage et chaque application a souvent un mélange d'entrées et sorties aléatoires et séquentielles.
- **VDI** : la VDI est l'un des principaux perturbateurs sur le marché du stockage. Les besoins de stockage de la VDI varient. Il ne faut parfois que 10 à 20 IOPS par utilisateur. À d'autres moments, par exemple, quand vous gérez des boot storms ou login storms (pics d'entrées-sorties de démarrage ou de connexion), les besoins en IOPS peuvent grimper en flèche.

Au fil des années, l'industrie a combiné toutes ces charges de travail variées. En d'autres termes, ce sont en fait leurs efforts de consolidation qui ont créé un monstre de stockage. Beaucoup d'environnements de stockage SAN (réseau de stockage) connaissent de gros problèmes en raison de l'effet I/O blender :

- ✓ La consolidation continue de VM contribue à des charges de travail aléatoires d'entrées et sorties, chacune avec un schéma individuel de lecture et d'écriture de données sur le support de stockage.
- ✓ Des flux I/O très aléatoires ont un effet délétère sur les performances générales car les VM sont en concurrence pour accéder aux ressources du disque.

Boot storm de la VDI

Le *boot storm de la VDI* est une situation qui traduit parfaitement un phénomène relativement nouveau en matière de stockage. Il a lieu quand un grand nombre d'utilisateurs essaient de démarrer leur machine virtuelle en même temps. Résultat : les périphériques de stockage n'arrivent plus à fournir en raison du nombre impressionnant de demandes.

C'est toutefois le début de la journée qui est la mort réelle du stockage. Alors que les ordinateurs démarrent, le système d'exploitation doit lire une tonne de données et la transférer à la mémoire pour que le système puisse être utilisé. Alors imaginez maintenant ce qui se passe quand des centaines ou des milliers d'utilisateurs démarrent en même temps leurs bureaux virtuels. Les systèmes de stockage existants s'écroulent sous le poids des I/O et les utilisateurs finissent parfois par devoir attendre longtemps avant que leurs systèmes aient entièrement démarré.



Ce problème est largement atténué par l'utilisation d'un stockage SSD comme couche de cache. L'ajout de ce type de service sans prendre en compte les lacunes administratives que cela introduit s'avère être une procédure opérationnelle standard depuis un certain temps et c'est une des raisons qui poussent certains à mettre en œuvre des îlots de ressources alors qu'ils veulent établir une VDI.

Atténuation

Les administrateurs des environnements existants peuvent essayer de nombreuses façons de résoudre les graves problèmes d'entrées et sorties qui surviennent avec des charges de travail partagées. En voici quelques-unes :



Lecture/Écriture des I/O

Dans les environnements traditionnels de datacenters avec un stockage partagé, les différences de performances entre la lecture et l'écriture de données sont incroyables. La lecture est généralement rapide et peut se faire en une seule opération I/O. Il en va différemment de l'écriture de données : elle peut demander jusqu'à six opérations I/O.

En passant de RAID 5 à RAID 6 pour améliorer la protection des données,

les administrateurs ont introduit un coût supplémentaire à l'équation du stockage. Une opération d'écriture RAID 6 demande la réalisation de six I/O. En effet, ce n'est pas seulement les données qui doivent être écrites, mais aussi les informations de parité qui doivent être écrites plusieurs fois dans RAID 6. Les calculs RAID ont aussi tendance à demander beaucoup de temps système pour réaliser les calculs de parité nécessaires à la protection des données.

- ✔ Acheter un environnement séparé pour soutenir chaque application.
- ✔ Acheter des périphériques de stockage autonomes sophistiqués qui comprennent des caractéristiques de hiérarchisation automatique.
- ✔ Acheter des niveaux multiples de stockage et les gérer séparément.

Qu'ont ces techniques d'atténuation en commun ? Elles exigent toutes des administrateurs qu'ils surprovisionnent le stockage, ce qui demande un investissement plus important en matériel physique. Elles exigent aussi plus de temps de configuration et de gestion par l'administrateur. À force, ces modèles peuvent devenir ingérables.

Points de contact multiples

Dans un environnement virtuel, les nombreux points de contact avec les données sont loin d'être idéaux. Réfléchissez au scénario suivant : un datacenter existant, mais fortement virtualisé a de nombreux serveurs VMware vSphere connectés à un réseau SAN. Le réseau SAN a des mécanismes de déduplication des données. L'entreprise utilise la méthode disk-to-disk-to-tape pour sauvegarder ses données et tous les jours, elle fait aussi une copie de certaines VM sur un datacenter externe. De cette manière, l'entreprise conserve des sauvegardes locales et obtient des capacités de reprise après sinistre.

Ce scénario intègre un degré important de redondance. La figure 2-1 examine le chemin parcouru par les données alors qu'elles passent à travers les divers processus associés à ce scénario.

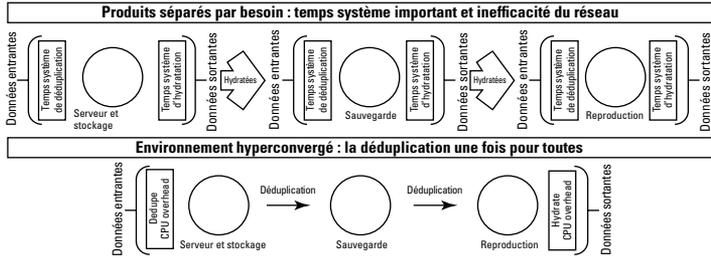


Figure 2-1 : l'infrastructure hyperconvergée demande moins de temps système et de largeur de bande de réseau que les systèmes non convergés.

Chaque fois qu'il faut hydrater des données puis les re-dédupliquer à mesure qu'elles progressent d'un composant à un autre, cela demande du temps système. La déduplication peut être une opération coûteuse et le traitement permanent des données dans des lieux différents a plusieurs inconvénients :

- L'utilisation continue du processeur pour traiter des données à plusieurs reprises limite le nombre de VM que l'environnement peut prendre en charge.
- Il faut plus de temps aux données pour traverser le réseau parce qu'elles n'ont pas été réduites alors qu'elles passent d'un service à un autre.
- Les coûts en largeur de bande WAN sont importants lorsque que les données non réduites passent par des connexions de réseaux étendus.

Ce n'est pas tout ! De nombreux systèmes de stockage y compris ceux liés à la protection de données emploient une méthode de déduplication *post-process*, plutôt qu'un processus de déduplication dont on dit qu'il est *à la volée*. La déduplication *post-process* signifie que les données ne sont dédupliquées qu'après qu'elles ont été écrites sur le disque. Voici les étapes :

1. **Écriture sur le disque des données non dédupliquées.** Il faut de la capacité et cela prend des IOPS.
2. **Relecture ultérieure depuis le disque.** Cela demande un nouveau point de contact avec les données, encore par le moteur de déduplication *post-process*, ce qui consomme encore plus d'IOPS et de temps système.

3. Recours au processeur pour dédupliquer ou compresser.

Une fois lues, les données doivent alors être à nouveau traitées, ce qui demande encore plus de temps système.

Cela veut dire que les données seront reproduites avant la déduplication, puis tout ce travail de déduplication pour gagner en capacité devra se faire sur les sites primaires et de reprise après sinistre. Cela consomme des ressources supplémentaires en termes de processeur, IOPS, capacité et largeur de bande de réseau. La déduplication post-process a recours à toutes ces ressources pour obtenir une réduction de la capacité de stockage. L'un ne compense pas l'autre.

Résultat : plus de coûts et moins d'efficacité. La meilleure issue dans un environnement est d'éviter l'écriture sur le disque. Dans un environnement hyperconvergé, en raison du cache de la mémoire vive (RAM), beaucoup d'opérations ne s'approchent jamais du stockage.

Dans le datacenter moderne, la gestion efficace des données est liée aux IOPS et à la largeur de bande WAN et non pas à la capacité de stockage. On ne manque plus de capacité, les disques des fournisseurs étant toujours plus puissants (6 To et plus !). À l'origine, les technologies de gestion efficace des données se concentraient sur le marché de la sauvegarde. L'objectif était de fournir une alternative à la technologie de bande. Pour que ce soit rentable, l'objectif primaire était d'héberger davantage de données sur le disque tout en assurant le débit nécessaire pour les sauvegardes. En bref, il s'agissait d'entasser 10 kg de données dans un sac de 2,5 kg. À l'époque, c'était la bonne solution.

Les disques durs ont fait peu de progrès alors que la capacité de stockage a augmenté. Les organisations n'ont pas de problèmes de capacité : leur problème tient aux IOPS, et se manifeste par de mauvaises performances. Avec l'ajout de la reprise après sinistre dans la plupart des environnements de clients, la demande de largeur de bande WAN a augmenté et là aussi, nous nous confrontons à un défi. Les technologies de réduction de données, comme la déduplication, ont pour but de faire face aux nouveaux enjeux de ressources, dont les besoins en largeur de bande WAN.

Cette situation étant établie, dans un environnement primaire de stockage, l'infrastructure doit être optimale au niveau des performances et de la latence, non pas à celui de la capacité et du débit. Cela demande une nouvelle technologie et une approche de la gestion efficace des données qui est systémique, une caractéristique de base de l'infrastructure hyperconvergée.

La déduplication à la volée fournit le niveau d'efficacité nécessaire et se fait en deux étapes : le traitement des données puis leur écriture.

La déduplication à la volée n'a recours au processeur qu'une seule fois, ce qui permet la réduction des IOPS, de la largeur de bande WAN et de la capacité de stockage. Ces ressources sont critiques, mais elles ne sont conservées que si la gestion efficace des données se fait à la volée.

Dans le datacenter moderne, la gestion efficace des données se résume à la mobilité et à la protection des données, en particulier quand il est question de sauvegarde et de restauration des données en ligne. La gestion efficace des données économise toutes les I/O autrefois nécessaires pour effectuer les opérations de sauvegarde et de restauration.

Décalage avec les politiques

Outre les difficultés de performances du monde post-virtualisation, les organisations virtualisées sont confrontées aux enjeux des politiques autant dans le monde physique que virtuel.

- **Physique** : Les serveurs physiques ont un mappage direct de l'application au serveur, à la baie de stockage, au LUN et à la politique de stockage. Ainsi, on a un environnement où la politique d'application est directement liée à une structure interne de la baie de stockage. Il n'y a pas d'abstraction. C'est cette approche qui rend les mises à niveau du stockage tellement complexes. Par exemple, une politique de reproduction est appliquée à un LUN dans une baie de stockage X à l'adresse IP Y et ordonne à ce LUN la reproduction sur la baie de stockage A à l'adresse IP B. Imaginez combien le remplacement d'une baie est complexe quand il y a deux baies à deux endroits et que les politiques de reproduction sont toutes entremêlées. Rien de surprenant à ce qu'il y ait tant d'administrateurs de systèmes de stockage en informatique.
- **Virtuel** : Dans le monde virtualisé, beaucoup d'applications sont sur un serveur et beaucoup de serveurs sur un seul LUN. Il n'est pas efficace d'appliquer une politique à un seul LUN si ce LUN représente les données de nombreuses applications (et serveurs). Dans un environnement hyperconvergé, les politiques de sauvegarde et de reproduction sont appliquées directement aux applications individuelles (ou VM). Pas besoin de gérer des LUN ou des ensembles RAID. Les politiques de reproduction indiquent une destination dans ce cas, un datacenter qui est prélevée de l'infrastructure. Cela permet à un administrateur de réaliser la mise à niveau d'une plateforme sans reconfiguration de la politique ni migration de données, ce qui accroît l'efficacité et diminue le risque.

Chapitre 3

Bienvenue dans le software-defined datacenter (SDDC)

Dans ce chapitre

- ▶ Tout virtualiser
 - ▶ Introduction de l'automatisation
 - ▶ L'IT as a Service
-



quel concept ! Les logiciels du datacenter moderne.

Réfléchissez à la situation il y a seulement cinq ans. Les datacenters existants étaient principalement du matériel physique. Les entreprises de gestion de stockage créaient leurs propres puces et boîtiers qu'ils expédiaient à leurs clients. Les vendeurs de mise en réseau adoptaient une approche semblable, en créant des circuits et des baies individuels pour leurs produits. Bien que cette approche ne soit pas nécessairement mauvaise, le matériel physique qui en résultait était inflexible et la couche logicielle flexible jouait un rôle de second plan.

Dans ce chapitre, je présente la nouvelle norme de datacenter : le software-defined datacenter (SDDC), dans lequel la place centrale est donnée aux logiciels et non plus au matériel. J'examine en détail ces caractéristiques qui définissent les SDDC, comme la virtualisation, l'automatisation et l'utilisation de l'IT as a Service (ITaaS).

Virtualisation

Chaque SDDC emploie un degré élevé de virtualisation, qui va plus loin que la virtualisation des serveurs. L'aspirateur de la virtualisation avale tout : stockage, serveurs et même les services associés

comme les équilibreurs de charge, les dispositifs d'optimisation du réseau WAN et les moteurs de déduplication. Rien n'est épargné. On élimine ainsi les îlots de processeur, mémoire, stockage et ressources de mise en réseau qui sont traditionnellement enfermés dans un périphérique à usage unique, comme un périphérique de sauvegarde sur disque, et l'on crée un pool de ressources partagées pour les applications commerciales mais aussi celles de l'infrastructure.

La virtualisation prélève les composants matériels du datacenter et leur superpose une couche logicielle commune : la couche de virtualisation, qui gère le matériel sous-jacent. Le matériel peut être un mélange sans nom, mais ça n'a plus aucune importance, grâce à la couche de virtualisation. S'assurer que les applications s'exécutent comme prévu est la seule chose dont l'administrateur du datacenter doit s'inquiéter. La couche de virtualisation gère le plus dur.

Automatisation

Les conseils d'administration sont de plus en plus nombreux à exiger que les entreprises fassent plus avec moins. Un des moyens les plus rapides d'améliorer son efficacité (et réduire ses coûts) est d'automatiser autant que possible les opérations de routine.

Jusqu'à maintenant, les nombreuses architectures d'informatique existantes étaient si variées et complexes que l'automatisation se résumait à un rêve illusoire. Grâce au SDDC, le rêve se rapproche un peu plus de la réalité.

La normalisation de l'architecture du datacenter qu'imposent les logiciels rend possibles des degrés plus élevés d'automatisation. De plus, la couche logicielle est elle-même bourrée de facilitateurs de l'automatisation, comme les interfaces de programmation d'applications (API). Il est bien plus facile de parvenir à l'automatisation si l'on obtient ce type d'assistance.

IT as a Service

Quand les ressources sont prélevées du matériel et qu'un grand nombre de techniques d'automatisation sont en place, les entreprises découvrent souvent qu'elles peuvent traiter de nombreux services informatiques exactement comme ça : en services.

Comme elles le font avec tous les autres services, les entreprises qui ont recours à l'ITaaS ont certaines attentes :

- ✔ **Prévisibilité** : le service doit fonctionner d'une manière et à un coût prévisibles. Le SDDC permet cette conformité.
- ✔ **Évolutivité** : les besoins de l'entreprise aujourd'hui peuvent être différents demain et le datacenter ne peut pas être un facteur limitateur quand une expansion devient nécessaire. En fait, un datacenter devrait être *facilitateur* de l'expansion de l'entreprise.
- ✔ **Meilleure utilisation** : les entreprises escomptent un avantage maximal des services qu'ils achètent. Comme un SDDC hyper-convergé est bâti sur des composants communs qui éliminent les îlots de ressources traditionnellement piégés dans les infrastructures physiques, il est extrêmement facile de parvenir à des taux élevés d'utilisation.
- ✔ **Moins de personnel** : Avec un SDDC, une entreprise a besoin de moins de personnes pour faire fonctionner un datacenter. Pour une simple raison : le SDDC bannit les îlots de ressources traditionnels pour les remplacer par une matrice générée par les logiciels.

Avoir moins de personnel se traduit directement par une baisse des coûts. En fait, la recherche conduite par Avaya suggère qu'un SDDC efficace peut faire passer les frais de personnel de 40 % du coût total de possession à seulement 20 %.
- ✔ **Temps de provisionnement réduit** : quand une entreprise investit dans le SDDC, elle espère des avantages commerciaux. Le SDDC apporte agilité et flexibilité, ce qui réduit le temps de provisionnement des nouveaux services dont les unités commerciales ont besoin.



Le matériel dans un monde de logiciels

Quand les gens entendent le terme *software-defined datacenter*, leur première question porte normalement sur l'endroit où les logiciels du SDDC sont censés s'exécuter. La réponse est simple : la couche logicielle s'exécute sur le matériel.

Mais si le SDDC est axé sur les logiciels, pourquoi a-t-on encore besoin de matériel ? Encore une fois, la réponse est simple : il est impossible d'exécuter un SDDC sans matériel.

Dans la plupart des cas, le matériel des SDDC ne ressemble pas beaucoup au matériel des environnements traditionnels. Alors que les datacenters existants ont du matériel propriétaire pour

gérer toute une myriade de périphériques, un SDDC utilise généralement un matériel de base.

Si un SDDC contient un matériel propriétaire, les logiciels s'en servent pour réaliser d'importantes fonctions. Dans le monde de l'hyperconvergence, ce type de matériel devient essentiellement partie intégrante des opérations standards du datacenter. Parce qu'il s'agit d'un matériel identique (qui n'est donc pas unique à chaque périphérique), il évolue en même temps que de nouvelles machines sont ajoutées au datacenter. Les logiciels restent maîtres dans cet environnement, mais rien ne se passerait sans le matériel

Chapitre 4

Ce que les entreprises attendent de l'informatique

Dans ce chapitre

- ▶ Gagner en efficacité
- ▶ Limiter les risques
- ▶ Devenir et rester agile

Saviez-vous que la seule raison d'être d'un service informatique n'est pas de jouer avec les technologies ? Étonnant, non ? Apparemment, il serait bien plus important que ce service toujours plus primordial détourne son attention du matériel pour la porter davantage sur l'activité commerciale.

Ce changement de focalisation n'est pas juste une bonne idée : c'est une tendance que les PDG et cadres dirigeants d'unités commerciales encouragent vivement car ils ont d'importants besoins à satisfaire. Les professionnels du secteur de la technologie qui veulent garder une longueur d'avance doivent affûter leur technique commerciale.

Les entreprises attendent des rendements sur les investissements importants en datacenters toujours plus élevés et elles sont de moins en moins prêtes à supporter les risques. Elles attendent d'un datacenter :

- ✓ qu'il améliore l'efficacité opérationnelle
- ✓ qu'il réduise les risques commerciaux
- ✓ qu'il soit suffisamment flexible et agile pour répondre à l'évolution des besoins commerciaux

Ce chapitre examine ces caractéristiques.

Plus d'efficacité

Votre patron est-il déjà entré dans votre bureau pour vous dire quelque chose de ce style ?

« Bob, il faut que nous parlions de votre bilan de performances. Vous êtes tout simplement trop efficace et il faut recadrer les choses de quelques degrés en-dessous. Si vous pouviez le faire samedi, ce serait suuuper ! »

Non ? C'est bien ce que je pensais. En fait c'est plutôt l'inverse : la pression est toujours plus grande sur les services IT pour qu'ils améliorent l'efficacité. Cela veut généralement dire la façon dont l'informatique fonctionne, des changements qui impliquent tout : de petites corrections de cap jusqu'aux initiatives majeures.



Un des plus gros avantages de l'architecture hyperconvergée est le fait qu'elle engendre plus d'efficacité tout en simplifiant les opérations.

Meilleure utilisation du temps

Comme le disait le poète Delmore Schwartz : « Le temps est le feu qui nous consume. » Il n'y a jamais eu de mots plus justes pour ceux qui triment chaque jour sur des tâches banales et répétitives. Quand on parle affaires, tout temps gaspillé à faire un travail banal est du temps brûlé : du temps qui aurait pu être utilisé à faire progresser les objectifs de l'entreprise.

La direction attend du service informatique qu'il utilise son temps à bon escient. Les opérations traditionnelles d'informatique ne suffisent plus. Les processus prolongés d'évaluation et d'intégration de produits ou les mesures étendues de retour sur investissement non plus. Le service informatique doit être plus rapide et rationnel que jamais.

Mise en correspondance des compétences et des tâches

Prenez un instant de recul pour réfléchir à ce que doivent véritablement gérer les informaticiens au quotidien : serveurs, hyperviseurs, périphériques de stockage, accélérateurs de réseau, logiciels de sauvegarde, périphériques de sauvegarde, technologie de reproduction et bien plus encore. Oubliez un moment les effets physiques de tout cet équipement sur le datacenter. Considérez plutôt le coût humain. Chacun de ces périphériques a une console d'administration

séparée que les opérateurs doivent savoir faire fonctionner. Par ailleurs, ne nous voilons pas la face, ces périphériques ne fonctionnent pas tous facilement les uns avec les autres.

Le fonctionnement de chaque périphérique demande un ensemble de compétences tout à fait unique et chaque compétence implique une formation continue. Même s'il est possible d'assurer la formation de quelques personnes du service informatique sur l'intégralité du datacenter, à un moment donné, ces personnes pourront passer à autre chose, et vous aurez peut-être du mal à trouver d'autres salariés qui disposent des mêmes compétences.

De plus, chaque fois que vous ajoutez une nouvelle ressource à l'environnement, il vous faut quelqu'un pour la gérer. Et si cette ressource évolue, vous aurez peut-être besoin de plus de personnel pour réussir à gérer la charge de travail. En soi, vous créez des îlots de ressources à mesure que vous allez de l'avant.



Les îlots de ressources sont intrinsèquement inefficaces. Plus l'environnement informatique est vaste, plus il est facile de réaliser des économies d'échelle sur l'exploitation.

En bref : les informaticiens sont écrasés sous le poids de l'infrastructure existante. Chaque ressource individuelle demande des compétences uniques et les entreprises n'embauchent pas leurs informaticiens au rythme auquel progressent les besoins techniques.

Bonne gestion des ressources

Les lois de la physique sont des forces immuables dans le paysage de l'informatique et ces lois naturelles deviennent apparentes quand on se promène autour d'un datacenter :

- Vous vous apercevez que deux objets ne peuvent pas occuper le même espace en même temps, ce qui explique pourquoi l'équipement se trouve sur des baies séparées.
- Vous vous apercevez des lois de la thermodynamique quand vous passez derrière une baie et sentez une bouffée d'air chaud, puis que vous passez sous l'appareil de refroidissement et ressentez un courant froid.
- Pour finir, vous découvrez l'électromagnétisme quand vous observez ce matériel gourmand en électricité en plein fonctionnement. (D'un point de vue positif, tous les voyants LED qui clignotent sont un spectacle lumineux assez impressionnant.)

Toutes ces ressources physiques (espace, puissance et refroidissement) ont besoin d'argent pour continuer à fonctionner. Chaque élément d'équipement ajouté à un datacenter aura besoin de toutes ces ressources.



Continuer d'ajouter du matériel sans prendre en compte votre utilisation des ressources ne servira d'aucune façon à améliorer votre efficacité générale.

Respect des budgets

Réfléchissez à votre premier projet de virtualisation. Pourquoi l'avez-vous entrepris ? Je serais prêt à parier que vous vouliez rationaliser la façon dont vous utilisez vos ressources et améliorer votre taux moyen d'utilisation des serveurs qui était normalement de 15 %.

Si vous n'utilisez pas tout votre matériel à un niveau raisonnable, vous jetez de l'argent qui pourrait vous servir à financer d'autres projets et à innover. Vous ne tirez peut-être pas le meilleur rendement possible d'un investissement incroyablement coûteux.



Votre budget IT peut être plus rationnel et mieux servir l'entreprise si vous repensez la façon dont vous fournissez des services de datacenter. Ne réfléchissez pas à chaque ressource individuelle comme à un îlot. Votre réflexion doit plutôt se faire à un niveau supérieur. Au lieu de vous concentrer sur des ressources individuelles, portez votre attention sur l'échelle globale de toutes les ressources comme votre procédure d'exploitation standard.

Diminution des risques

Des risques peuvent s'insinuer dans un système informatique autrement sans failles, à plusieurs endroits.

✓ **Approvisionnement** : il peut être difficile de garder l'œil sur tous les petits détails quand il y a tant de matériel à entretenir dans le datacenter. Avant d'acheter un nouvel équipement, posez-vous des questions comme celles qui suivent.

- Êtes-vous sûr que le réseau SAN sélectionné a suffisamment de capacité en termes de téraoctets *et* de performances en termes d'IOPS (opérations d'entrée-sortie par seconde) ?
- Si vous agrandissez un système actuel, son expansion créera-t-elle un risque de temps d'arrêt ou de perte de données ?

- Si vous mettez à jour une baie de stockage existante, les révisions des firmwares des composants sont-elles toutes les mêmes ou prises en charge par le nouveau matériel ?

✓ **Opérations** : de façon générale, il faut qu'un datacenter ait un double de chaque chose si l'on veut garder les niveaux de disponibilité que l'entreprise attend. Autrement, on court le risque de pannes d'alimentation que la direction tend à détester.

En informatique, la redondance est la norme, mais c'est une norme coûteuse. De plus, elle demande du personnel spécialisé pour que les différentes fonctionnalités associées à chaque produit restent tout à fait disponibles.

✓ **Protection des données** : trop d'entreprises ne planifient pas soigneusement leurs mécanismes de protection des données ou s'appuient sur des services multiples de nombreuses sociétés. Par conséquent, on assiste à un jeu de faute rejetée si quelque chose d'inattendu se produit. Quand la priorité absolue est la restauration, personne ne tient à ce que les fournisseurs se disputent pour savoir qui est responsable. N'oubliez pas : en matière de protection des données, il n'est pas question de sauvegarde, mais de *restauration*.



Les directeurs des systèmes d'information et le personnel du service informatique veulent, et doivent, diminuer les risques pour leur organisation. Les systèmes et les applications doivent rester tout à fait disponibles et les données doivent être sécurisées. Malheureusement, à mesure que des éléments divers de matériel sont installés, l'atteinte de ces objectifs cruciaux devient plus difficile.



Les entreprises peuvent réduire les risques en adoptant une architecture hyperconvergée. Les systèmes hyperconvergés comprennent tous les composants nécessaires pour faire fonctionner un datacenter sans la complexité des solutions existantes.

Plus d'agilité

Mettre rapidement en place un produit ou un service apporte souvent des avantages commerciaux à long termes. Les marchés actuels sont féroces et on ne peut pas laisser des problèmes informatiques entraver des opérations commerciales critiques. C'est une des raisons qui rendent le cloud public aussi attrayant aux yeux des usagers professionnels.

Face à un délai à répondre à ses besoins, l'entreprise aura tendance à prendre les choses en main à sa manière, c'est ce qu'on appelle

la « shadow IT » et cela arrive plus souvent qu'on ne le croit. Si un élément constitutif de l'entreprise fait une demande de fonction ou de service à laquelle le service informatique ne peut pas répondre pleinement, alors il se procurera le service dans le cloud, en donnant simplement son numéro de carte de crédit, sans que le service informatique soit impliqué, ni même informé de son existence. Les pratiques de la shadow IT (informatique fantôme) sont une source potentielle de risques.

Pour répondre aux unités commerciales qui demandent plus de rapidité et d'agilité, de nombreux services informatiques construisent simplement des infrastructures toujours plus larges, qui manquent de flexibilité et sont difficiles à gérer. Un tel système finit par se transformer en château de cartes et le plus petit problème peut tout faire écrouler. Ce scénario ne correspond guère à une infrastructure d'informatique agile.



Tout informaticien devrait adopter deux mantras : *agilité* et *rapidité*. Il est fondamental de s'assurer que l'infrastructure soit agile pour que le service informatique puisse facilement et rapidement déployer de nouvelles applications et de nouveaux services pour satisfaire les demandes de l'entreprise.

Chapitre 5

Comment le cloud fait évoluer l'informatique

Dans ce chapitre

- ▶ Emploi de ressources évolutives partagées
- ▶ Accent sur l'automatisation
- ▶ Gestion judicieuse des achats

La virtualisation est juste une des tendances importantes ayant une incidence sur l'informatique. Il est difficile de nier que le cloud, lui aussi, affecte l'informatique. Points de réflexion :

- ✔ Les unités commerciales et les services informatiques peuvent obtenir des services simplement avec une carte de crédit.
- ✔ Les principaux prestataires de service dans le cloud comme Google et Facebook changent les attentes quant au fonctionnement d'un datacenter, car leurs environnements gigantesques n'ont rien à voir avec les datacenters existants. Même si la plupart des organisations n'ont pas besoin d'une telle échelle, les meilleurs éléments de la conception architecturale de ces clouds ont été rapportés au monde hyperconvergé et présentés dans une offre abordable.

En matière d'infrastructure hyperconvergée, seule la seconde tendance a réellement de la pertinence et c'est le sujet de ce chapitre.

Évolutivité et partage des ressources

Ce qui distingue les environnements Google et Facebook, entre autres choses, ce sont leur impressionnante évolutivité et leur coût raisonnable. Un grand nombre des principes de ces clouds ont été adaptés pour des environnements plus petits et proposés sous forme de produits hyperconvergés que n'importe quelle entreprise peut acheter.

Conception axée sur les logiciels

Comme je l'ai mentionné au chapitre 3, dans un datacenter, la domination des logiciels sur le matériel peut porter de très bons fruits. Il y a des années que les entreprises comme Google ont découvert ce potentiel et qu'elles ont dompté leur matériel en l'enveloppant au milieu de couches logicielles. Un fichier de données Google est géré par le système logiciel global de gestion des fichiers que l'entreprise diffuse très largement. Ce système de gestion de fichiers ne s'intéresse pas au matériel sous-jacent. Il respecte simplement les règles intégrées à la couche logicielle qui veille à la bonne sauvegarde du fichier et aux bons niveaux de protection des données. Même avec l'expansion de l'infrastructure de Google, l'administrateur ne s'inquiète pas du lieu où ce fichier réside.

Économies d'échelle

Dans un environnement de datacenter existant, il peut être coûteux d'agrandir l'environnement en raison de la nature propre à chaque élément du matériel. Plus l'environnement est varié, plus il est difficile d'assurer sa maintenance.

Matériel générique

Les entreprises comme Google et Facebook adaptent leurs environnements sans dépendre de composants de marque coûteux. Elles utilisent plutôt du matériel générique. Pour certains, s'il est associé à l'environnement du datacenter, le terme *générique* est synonyme de *bon marché* ou de *peu fiable*. Vous voulez que je vous dise ? Ils ont raison jusqu'à un certain point.



Toutefois, quand on considère le rôle du matériel générique dans un environnement hyperconvergé, il faut garder à l'esprit que le matériel vient en second plan des logiciels. Dans cet environnement, la couche logicielle est créée en sachant que le matériel peut (et finira par) échouer. L'architecture logicielle est conçue pour anticiper et gérer les pannes éventuelles de matériel. Le matériel générique n'est pas bon marché, mais il est beaucoup moins cher que le matériel propriétaire. Il est également interchangeable avec d'autres éléments. Un fournisseur d'hyperconvergence peut changer de plate-forme matérielle sans avoir à recoder tout son système. Grâce au matériel générique, les fournisseurs d'hyperconvergence garantissent à leurs clients l'accès à un matériel abordable sans interruption, puisque les modifications sont simples et rapides.

Évolutivité granulaire

Réfléchissez à la façon dont vous vous procurez votre technologie de datacenter à l'heure actuelle, en particulier quand il est question

de stockage et d'équipement autre que le serveur. Pour le cycle de vie prévu de cet équipement, vous achetez probablement toute la puissance et la capacité dont vous aurez besoin, avec juste un peu plus de capacité « au cas où ».

Combien de temps vous faudra-t-il pour utiliser toute la capacité que vous avez déjà acquise ? Peut-être que vous ne l'utiliserez jamais. Quel gaspillage ! Mais d'un autre côté, vous trouverez peut-être que vous devez agrandir votre environnement plus tôt que prévu. Les entreprises dans le cloud ne créent pas de plans complexes de mise à niveau d'infrastructures à chaque fois qu'elles s'agrandissent. Elles ajoutent simplement des unités standardisées d'infrastructure à leur environnement. C'est leur modèle d'échelle : tout est question de pouvoir passer au niveau supérieur d'infrastructure par petits échelons, selon les besoins.

Flexibilité des ressources

L'infrastructure hyperconvergée adopte une approche granulaire vis-à-vis de l'évolutivité du datacenter. Les clients n'ont plus besoin d'agrandir un seul composant ou une seule baie de matériel à la fois ; ils peuvent simplement ajouter un autre nœud de périphérique à un environnement homogène. Tout l'environnement est un pool énorme de ressources virtualisées. Selon les besoins, les clients peuvent agrandir ce pool, rapidement et facilement, d'une façon logique d'un point de vue économique.

Rendre l'automatisation possible

Croyez-vous que le prestataire cloud typique passe une quantité anormale de temps à recréer des politiques et des procédures individuelles chaque fois qu'il fait des changements ou ajoute un équipement à son datacenter ? Bien sûr que non. Dans le cloud, le changement est constant, alors il est crucial de s'assurer que les changements se font sans perturbation. Dans le monde de l'informatique d'entreprise, les choses devraient fonctionner de la même manière. Un changement de matériel dans le datacenter ne devrait pas impliquer une reconfiguration de toutes vos machines virtuelles et politiques.

Retirer la politique de l'infrastructure

Puisque le matériel n'est pas au centre du software-defined datacenter (SDDC ; voir le chapitre 3), pourquoi voudriez-vous écrire des politiques qui visent des périphériques matériels particuliers ? De plus, comme l'entreprise utilise à la base des machines virtuelles (VM) pour ses charges de travail, pourquoi une VM serait-elle tenue

de respecter des politiques liées à des composants de l'infrastructure sous-jacente ?

Imaginez un scénario selon lequel vous définissez des politiques qui font passer les charges de travail d'un LUN particulier à un autre pour des raisons de reproduction. Maintenant, multipliez cette politique par 1 000. Le jour où vous devrez remplacer un LUN, vous vous retrouvez avec un entremêlement de politiques de LUN à LUN. Vous devez trouver chaque politique individuelle et la reconfigurer pour qu'elle s'oriente sur le nouveau matériel.

Les politiques devraient être d'une nature beaucoup plus générale, pour permettre à l'infrastructure de prendre des décisions granulaires. Au lieu, par exemple, de se montrer spécifique dans une politique de LUN, les politiques devraient être aussi simples que : « Reproduire la VM à San Francisco ».

Pourquoi est-ce une bonne chose ? Une politique aussi générale permet d'effectuer une remise à niveau complète de la technologie à San Francisco sans migration de données ni reconfiguration de politiques. Comme la politique est définie au niveau du datacenter, toutes les politiques appliquées au trafic entrant et sortant restent en place. Fabuleux !

Adoption d'une approche centrée sur les VM

Dans le cloud, la charge de travail a le premier rôle. Dans le cas de l'informatique d'entreprise, ces charges de travail sont des VM individuelles. Quand il est question de politiques dans les environnements cloud, la VM est le centre du monde. Tout est lié à l'application de politiques aux VM (et non pas à des LUN, partages, datastores ou d'autres structures). Ayez une pensée pour l'administrateur de VM qui est axé sur les VM. Pourquoi l'administrateur n'assignerait-il pas des politiques de sauvegarde, qualité de service et reproduction à une VM ?

En informatique, le besoin d'appliquer des politiques à travers des domaines de ressources individuelles engendre des problèmes de fonctionnement fondamentaux. Dans le cloud et en matière d'hyperconvergence, les politiques sont simplement appelées politiques. Il n'y a pas de *politiques* pour les LUN, pour les caches, pour la reproduction, etc., juste des politiques.

Comprendre l'économie du cloud

Les prestataires cloud et les organisations d'informatique pour les entreprises s'appuient sur des modèles économiques très différents pour le fonctionnement de leurs environnements. Les directeurs des

systèmes d'information attendent de l'infrastructure informatique de leur entreprise qu'elle dure de nombreuses années, alors ils achètent suffisamment de capacité et de performances pour une telle durée. Dans de nombreux cas cependant, tout le potentiel de l'infrastructure achetée n'est jamais réalisé. Les directeurs des systèmes d'information surachètent pour s'assurer qu'ils ont la capacité nécessaire pour durer pendant tout le cycle de vie.

À l'inverse, ils ont souvent tendance, à dessein ou par inadvertance, à sous-acheter de l'infrastructure. L'organisation doit alors acheter de nouvelles ressources individuelles quand elle commence à en manquer. Ainsi, il faut sans cesse surveiller des ressources individuelles et réagir si nécessaire en espérant que le produit que vous avez actuellement n'est pas arrivé à sa fin de vie.

Maintenant, réfléchissez aux fournisseurs cloud qui ne font pas un achat énorme tous les cinq ans. Le faire serait de la folie pour plusieurs raisons :

- ✓ *Beaucoup* de matériel devrait être acheté d'avance.
- ✓ Planifier exactement trois à cinq années de besoins en termes de ressources dans ces types d'environnement est sans doute impossible.

Au lieu de cela, les organisations cloud achètent au rythme de leur croissance. L'échelle opérationnelle et l'homogénéité sont des éléments fondamentaux de l'ADN des prestataires cloud, alors ils ajoutent simplement davantage de ressources standard quand ils en ont besoin.

Le cloud public est très attrayant pour l'entreprise. Ce service instantané est élastique et coûte quelques centimes par heure. Que pourrait-il y avoir de mieux ? Il ne convient pas à toutes les applications : il a de nombreux inconvénients, en particulier au niveau de la prévision des coûts. Le coût réel du cloud public augmente fortement par rapport au coût d'une performance prévisible de stockage, de grande disponibilité, de sauvegarde, de reprise après sinistre, de mise en réseau privé, pour n'en citer que quelques-uns. Le service informatique finit par payer pour un serveur qui fonctionne à 15 % de sa capacité et le prestataire cloud a l'avantage de rassembler ces VM sur un seul hôte.

Les applications informatiques sont conçues pour une infrastructure très disponible, de reprise après sinistre, de sauvegarde, de restauration et d'autres services nécessaires (ce qui explique que l'IT interne ne ressemble pas à Facebook ou à Google). L'ensemble des exigences que les applications informatiques ont de l'infrastructure

est différent. Ainsi, une infrastructure hyperconvergée *doit* satisfaire ces exigences.

Les unités commerciales peuvent ne pas comprendre ces nuances et peuvent être forcées à acheter des services cloud sans comprendre les tenants et les aboutissants. La montée de la shadow IT (des services autres que le service informatique créent leurs propres systèmes, voir le chapitre 4) est réelle et le cloud est le parfait complice pour rendre possible cette tendance. La shadow IT existe toutefois soit parce que le service informatique n'est pas capable de fournir les types de services que les unités commerciales exigent, soit parce qu'ils ne sont pas assez réactifs. Alors ces services se tournent vers le cloud pour répondre à des besoins individuels, donnant lieu à des services fragmentés, à d'éventuelles violations de sécurité et généralement, à des problèmes de qualité des données.

L'hyperconvergence met à la disposition des services informatiques l'économie et la flexibilité des infrastructures de consommation cloud sans que les performances, la fiabilité ou la disponibilité soient compromises. Au lieu de faire d'énormes achats à quelques années d'intervalle, le service informatique ajoute au datacenter les blocs constituants d'infrastructure nécessaires à ce moment-là. Cette approche permet à l'entreprise d'atteindre le seuil de rentabilité beaucoup plus vite pour l'expansion de son environnement.

La figure 5-1 présente les systèmes intégrés (qui s'adaptent par grandes étapes ; voir le chapitre 6) et l'hyperconvergence (qui s'adapte par étapes granulaires ; voir les chapitres 6 et 7), en montrant les ressources gaspillées au-dessus du trait ondulé.

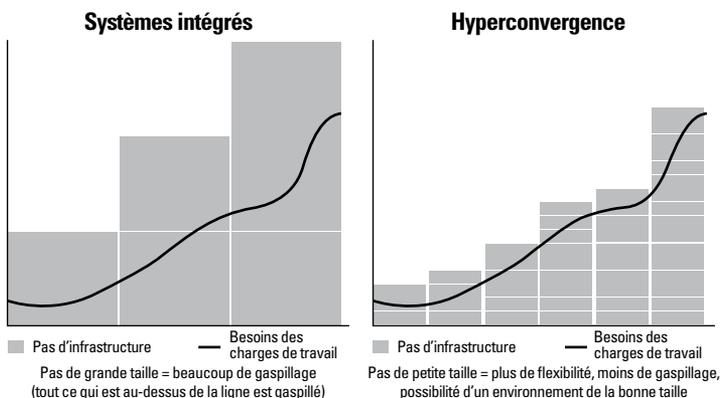


Figure 5-1 : L'unité d'échelle a un impact considérable sur l'économie.

Chapitre 6

Compréhension de l'infrastructure hyperconvergée

Dans ce chapitre

- ▶ L'évolution de la convergence
- ▶ Introduction de l'infrastructure hyperconvergée
- ▶ Tirer le maximum des différences

Le marché de l'infrastructure informatique subit une transformation sans précédent. La transformation la plus importante se reflète dans deux tendances majeures : la convergence et les software-defined datacenters (SDDC). Ces deux tendances répondent aux réalités de l'informatique : encombrement, complexité et coût élevé des infrastructures ; elles représentent des tentatives de simplification de l'informatique et de réduction du coût global de propriété d'une infrastructure. Les environnements actuels d'infrastructures comprennent généralement 8 à 12 produits matériels ou logiciels d'autant de fournisseurs, où chaque produit dispose d'une interface de gestion différente et demande une formation distincte.

Chaque produit dans ce type de baie existante est toutefois surprovisionné : il utilise ses propres ressources (processeur, DRAM, stockage, etc.) pour gérer les charges de travail de pics intermittents d'applications résidentes. La valeur d'un seul pool de ressources partagées, que permet la virtualisation du serveur, est encore limitée à la seule couche du serveur. Tous les autres produits sont des îlots de ressources surprovisionnées qui ne sont pas partagées. Ainsi, le faible taux d'utilisation de toute la baie a comme répercussion des coûts élevés d'acquisition, d'espace et de puissance. En termes simples, trop de ressources sont gaspillées dans les environnements existants actuels.

Ce chapitre explore une solution de pointe : la convergence qui mène en dernier lieu à l'hyperconvergence.

L'évolution de la convergence

Les parties qui suivent décrivent l'évolution de la convergence au cours des dernières années.

Systèmes intégrés

Les toutes premières solutions de convergence d'infrastructure ont des fonctionnalités complètes de réseau, calcul, stockage et virtualisation, mais dans de nombreux cas, elles ne sont que des conglomérations de matériels et logiciels existants, qui ne présentent que peu d'innovation réelle au niveau des caractéristiques produit qui peuvent être utilisées.

Des systèmes intégrés apportent quelques avantages. Le plus notable est que les clients bénéficient d'un point de contact unique pour leur infrastructure, de son achat jusqu'à sa fin de vie. Ces systèmes sont toujours testés et presque toujours livrés sur le site du client complètement contenus dans des baies et entièrement câblés, si bien qu'ils sont prêts à l'emploi.

L'inconvénient, c'est souvent le grand pas d'une taille à l'autre de ces systèmes. Quand vous avez besoin de plus de puissance, vous devez acheter un gros morceau d'infrastructure. Ces produits ne résolvent pas non plus toujours les graves problèmes que de nombreuses organisations rencontrent (voir le chapitre 2).

L'infrastructure convergée

Les produits de l'infrastructure convergée combinent les composants du serveur et de stockage en une seule machine, ce qui élimine le besoin d'un stockage SAN dédié.

Ces systèmes offrent une solution unique localisée de pool de ressources, permettant une gestion simplifiée et un déploiement plus rapide. Ils ont effectivement virtualisé la couche de stockage qui peut s'exécuter dans la plateforme de virtualisation. Le coût d'acquisition global est moins élevé et la gestion simplifiée (tout au moins du serveur et des ressources de stockage). Avec ces systèmes, l'utilisation globale des ressources est plus élevée qu'avec une infrastructure d'îlots existante.

L'infrastructure convergée a cependant certaines limitations :

- Les systèmes ne comprennent souvent que les composants du serveur et des ressources de stockage.
- Une grande partie des difficultés fondamentales de gestion des données n'a pas été résolue. C'est la fonctionnalité d'une baie

de stockage traditionnelle transférée dans la plateforme de virtualisation.

- ✔ Les rapports de ressources (processeur, stockage, réseau, etc.) sont fixes et offrent donc moins de flexibilité que ce dont certaines organisations ont besoin.
- ✔ Les produits ne peuvent pas toujours être utilisés par l'infrastructure actuelle. En d'autres termes, vous pourriez ne pas être en mesure d'utiliser le stockage d'une machine à infrastructure convergée depuis des systèmes existants existants. En fait, cela vous force à créer un îlot séparé de ressources.

Pour ces raisons, les systèmes d'infrastructure convergée ne corrigent pas suffisamment les problèmes de performances dans chaque infrastructure existante.

De la même façon, concernant les données, les systèmes ne résolvent pas toutes les difficultés liées aux données, car tous les périphériques de gestion efficace des données ne sont pas convergés. La gestion peut être améliorée, mais il ne s'agit pas d'une gestion globale unifiée.

L'infrastructure hyperconvergée

Apparaît ensuite l'hyperconvergence, connue aussi sous le nom d'*infrastructure hyperconvergée*, qui constitue l'étape logique suivante dans l'évolution de la convergence des infrastructures. L'hyperconvergence permet la simplification et des économies en consolidant toutes les fonctionnalités requises en une seule baie d'infrastructure qui s'exécute sur un pool efficace et élastique de ressources x86. L'architecture sous-jacente des données a été totalement réinventée, ce qui permet une simplification radicale de leur gestion. En conséquence, l'infrastructure hyperconvergée tient la promesse du SDDC sur le plan technique. Elle fait aussi progresser les avantages de la convergence, dont un pool unique de ressources partagées.

L'hyperconvergence ne s'arrête pas aux serveurs et au stockage, mais regroupe en son sein bon nombre d'autres caractéristiques qui rendent obsolètes les services existants dans certains environnements informatiques, comme ceux qui suivent :

- ✔ Produits de protection des données (sauvegarde, reproduction)
- ✔ Périphériques de déduplication
- ✔ Périphériques d'optimisation du réseau WAN

- ✓ Baies SSD
- ✓ Baies de cache SSD
- ✓ Passerelles vers le cloud public
- ✓ Périphériques ou logiciels de reproduction

Dans le chapitre 7, je me penche sur la façon dont l'hyperconvergence permet à la convergence de passer à l'étape supérieure dans le datacenter et dont elle offre une multitude d'avantages autant au service informatique qu'à l'entreprise.

Caractéristiques de la convergence

Les options de convergence qui précèdent s'appuient l'une sur les autres, mais ont chacune des différences fondamentales. La figure 6-1 décrit les caractéristiques majeures qui définissent chaque type de convergence.

	Attributs techniques			Avantages organisationnels	
	Efficacité des données	Pool unique de ressources partagées	Gestion globale	Amélioration du TCO	Simplification
Systèmes intégrés	<input type="checkbox"/>	<input type="checkbox"/> Pool de ressources limité à la couche des serveurs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Délais avant le déploiement et les gains administratifs
Convergence	<input type="checkbox"/>	<input checked="" type="checkbox"/> Limité au serveur primaire et aux ressources de stockage ; autres ressources non comprises	<input type="checkbox"/>	<input checked="" type="checkbox"/> Gains de TCO dus avant tout à la réduction du matériel existant ; ne traite pas la sauvegarde, la reproduction et la reprise après sinistre	<input checked="" type="checkbox"/> Moins de produits à gérer
Hyperconvergence	<input checked="" type="checkbox"/> L'architecture de gestion des données commence par réaliser une seule fois la déduplication, la compression et l'optimisation des données	<input checked="" type="checkbox"/> Toutes les ressources du datacenter sont réunies dans la baie des ressources	<input checked="" type="checkbox"/> Gestion complète de toutes les ressources de l'infrastructure et des machines virtuelles ; point unique d'administration	<input checked="" type="checkbox"/> Gains importants de TCO par la réduction des ressources matérielles, la simplification des opérations et l'automatisation	<input checked="" type="checkbox"/> Réduction du matériel éparpillé à travers les datacenters, facilité de gestion, orientations sur les VM

Figure 6-1 : comparaison des améliorations avec l'évolution de la convergence.

Chaque caractéristique est critique pour obtenir toutes les qualités qu'exigent les entreprises de l'informatique dans l'ère moderne : simple, économique et écologique.

Chapitre 7

Les neuf choses qu'une infrastructure hyperconvergée peut faire pour vous.

Dans ce chapitre

- ▶ Focalisation sur les logiciels
- ▶ Centralisation de la gestion
- ▶ Renforcement de la protection des données

Comment une infrastructure hyperconvergée réunit-elle toutes les tendances importantes que l'informatique d'entreprise a du mal à gérer ? Jetez un coup d'œil :

- ✔ L'hyperconvergence est la personnalisation du software-defined datacenter (SDDC ; voir le chapitre 3). Basée dans les logiciels, elle apporte la flexibilité et l'agilité que l'entreprise exige de son informatique.
- ✔ Les opérateurs de clouds ont bien imaginé leur modèle économique. L'hyperconvergence donne à l'informatique d'entreprise un modèle économique similaire à celui du cloud, qui permet d'atteindre le seuil de rentabilisation plus rapidement pour les dépenses de datacenter et un plus faible coût total de possession pour la solution complète. Une infrastructure hyperconvergée offre les avantages économiques du cloud tout en livrant les performances, la haute disponibilité et la fiabilité qu'exige l'entreprise.
- ✔ Le stockage flash élimine certains problèmes de performances, mais pas tous. Les options hyperconvergées comprennent des systèmes 100 % flash ou des systèmes qui combinent un stockage flash à un disque rotatif, pour que le service informatique

puisse mettre en équation la capacité et les performances avec les exigences de l'entreprise, en éliminant les îlots de ressources (voir le chapitre 2).

- ✓ Le marché de l'infrastructure convergée propose une démarche de fournisseur unique pour l'approvisionnement, la mise en œuvre et l'exploitation. Il n'y a aucun jeu de faute rejetée et juste un numéro d'appel quand un problème survient concernant le datacenter.

Dans ce chapitre, j'approfondis le sujet de l'hyperconvergence et vous présente dix façons de résoudre avec elle les défis inhérents à des datacenters virtualisés (voir le chapitre 2).

Focalisation sur les logiciels

L'hyperconvergence est la quintessence du SDDC (software-defined datacenter) abordé au chapitre 3. L'hyperconvergence étant basée sur des logiciels, elle donne la flexibilité nécessaire pour répondre aux besoins actuels et futurs de l'entreprise, sans devoir démonter et remplacer des composants de l'infrastructure. Encore mieux, à mesure que les fournisseurs ajoutent de nouvelles caractéristiques dans les nouvelles versions de logiciels, les clients profitent immédiatement des avantages de ces caractéristiques, sans devoir remplacer de matériel.

Systèmes et gestion centralisés

Dans une infrastructure hyperconvergée, tous les composants (calcul, stockage, sauvegarde sur disque, fonctionnalité de passerelle vers le cloud, etc.) sont combinés en un pool unique de ressources partagées grâce à la technologie de l'hyperviseur. Cette conception simple et efficace permet au service informatique de gérer des ressources regroupées à travers des nœuds individuels, en un seul système fédéré.

La centralisation et l'intégration de masse se produisent aussi au niveau de la gestion. Indépendamment de l'éparpillement éventuel des ressources physiques, les systèmes hyperconvergés les gèrent comme s'ils étaient installés côte à côte. Les ressources éparpillées entre de multiples datacenters physiques sont gérées depuis une interface centralisée unique. Toutes les fonctions de gestion du système et des données sont également traitées à partir de cette interface.

Plus d'agilité

L'agilité revêt beaucoup d'importance dans l'informatique moderne. L'entreprise attend de son service informatique qu'il réagisse rapidement dès que de nouveaux besoins surviennent, mais les environnements existants le forcent à recourir à des multitudes de ressources pour y parvenir. L'infrastructure hyperconvergée permet au service informatique d'obtenir beaucoup plus vite des résultats positifs.

Être agile c'est en partie être capable de déplacer des charges de travail selon les besoins. Dans un monde hyperconvergé, toutes les ressources dans tous les datacenters physiques résident sous une seule tutelle administrative (voir la partie qui précède). Dans de tels environnements, la migration des charges de travail est simple comme bonjour, en particulier dans une solution hyperconvergée dont l'offre propose comme élément de base une déduplication cohérente. Les données réduites sont bien plus faciles à exploiter que sous leur forme complètement développée et grâce à elles, l'informatique peut faire les choses plus vite.

Évolutivité et efficacité

L'hyperconvergence est une approche de constituants évolutifs qui permet l'expansion des systèmes informatiques par l'ajout d'unités, tout comme un jeu de LÉGO. L'évolutivité granulaire est une des caractéristiques de base de cette infrastructure. À l'inverse des systèmes intégrés, qui exigent souvent de gros investissements, le pas pour passer d'une solution hyperconvergée à celle d'une taille supérieure est beaucoup plus petit. *La taille du pas* est la quantité d'infrastructure que doit acheter une entreprise pour passer au niveau supérieur d'infrastructure. Plus le pas est grand, plus les montants à avancer seront importants.



Plus le pas est grand, plus il faudra de temps pour utiliser pleinement les nouvelles ressources qui ont été ajoutées dans le cadre de l'expansion. Une taille de pas plus petite permet une utilisation beaucoup plus rationnelle des ressources. À mesure que de nouvelles ressources sont nécessaires, il est facile d'ajouter des nœuds à une infrastructure hyperconvergée.

Ça ressemble aux Borgs, non (dans le bon sens) ?

Une infrastructure rentable

Les coûts d'entrée des systèmes hyperconvergés sont faibles, tout comme le coût total de propriété (TCO) comparés aux systèmes intégrés équivalents et à l'infrastructure existante. (Pour en savoir plus sur les systèmes intégrés, voir le chapitre 6.)

Automatisation facile

L'automatisation est un élément clé du SDDC (voir le chapitre 3) et il va de pair avec l'hyperconvergence. Quand toutes les ressources sont véritablement combinées et quand des outils de gestion centralisée sont en place, la fonctionnalité administrative comprend des possibilités de programmation, mais aussi des options de génération de script.

De plus, le service informatique n'a pas à essayer de créer des structures automatisées avec du matériel de fabricants ou des lignes de produits différents. Tout est encapsulé en un seul environnement bien organisé.

Focalisation sur les VM

La virtualisation est le fondement du SDDC (voir le chapitre 3). Les systèmes hyperconvergés ont recours à des machines virtuelles (VM) qui sont les constituants les plus basiques de l'environnement. Toutes les ressources (stockage, sauvegarde, reproduction, équilibrage de charge, etc.) prennent en charge des VM individuelles.

Par conséquent, les politiques dans l'environnement hyperconvergé s'organisent autour des VM, tout comme l'ensemble des options de gestion disponibles dans le système, comme les politiques de protection des données, qui sont souvent définies par des outils tiers dans des systèmes existants. Grâce à l'hyperconvergence, les politiques intégrées de protection des données et le contrôle ont lieu au niveau même des VM. (J'aborde la protection des données plus loin dans ce chapitre.)

Le centrage sur les VM est également apparent alors que les charges de travail doivent être déplacées vers des datacenters différents et entre des services, comme la sauvegarde et la reproduction. Pour l'administrateur, la machine virtuelle est toujours au centre des opérations, pas le datacenter ni les services sous-jacents, comme le stockage.

Ressources partagées

L'hyperconvergence permet aux organisations de déployer de nombreuses sortes d'applications dans un seul pool de ressources partagées sans s'inquiéter de l'effet I/O blender redouté (voir le chapitre 2), qui anéantit les performances des VM.

Comment l'hyperconvergence rend-elle possible ce type de déploiement ? Les systèmes hyperconvergés comprennent des types multiples de stockage (soit 100 % flash, soit une combinaison de stockage SSD et de disque rotatif) dans chaque appareil. Plusieurs téraoctets de chaque type de stockage peuvent être installés sur un seul appareil. Comme il faut plusieurs appareils pour obtenir la complète redondance de l'environnement et la protection des données, les deux types de stockage sont en quantité largement suffisante. Dans les systèmes hyperconvergés, la focalisation sur les VM permet aussi au système de voir à travers l'effet I/O blender et d'optimiser en fonction du profil des entrées et sorties de la VM individuelle.

Le stockage hyperconvergé gère adroitement les charges de travail aléatoires et séquentielles. Encore mieux, avec plusieurs périphériques de stockage SSD dans un cluster hyperconvergé, le nombre des IOPS (opérations d'entrée-sortie par seconde) est largement suffisant pour supporter mêmes les charges de travail les plus intensives — y compris les boot storms ou login storms de l'infrastructure de bureau virtuel (VDI) (voir le chapitre 2).

Le pool de ressources partagées permet aussi leur utilisation efficace, ce qui améliore les performances et la capacité, tout comme ces toutes premières initiatives de consolidation de serveurs que vous avez entreprises quand vous vous êtes initialement lancé sur la voie de la virtualisation. En chemin, cependant, vous pouvez avoir créé de nouveaux îlots en raison des difficultés post-virtualisation dont nous avons déjà parlé. Les îlots de ressources comportent les mêmes difficultés d'utilisation que vos anciens environnements physiques mentionnés. Grâce à l'hyperconvergence, vous pouvez vous libérer du besoin de créer des îlots de ressources juste pour répondre aux besoins d'entrées et de sorties d'applications particulières. L'environnement même gère toutes les attributions (processeur, RAM, capacité et IOPS) si bien que les administrateurs peuvent se concentrer sur l'application plutôt que sur les besoins individuels de ressources.

La bonne combinaison de structures informatiques sur site, basées sur une architecture définie par logiciel peut vous permettre de fournir à votre entreprise une prestation d'IT-as-a-service, tout en évitant les dangers du cloud public. L'entreprise profite de dépenses d'informatique plus faibles tout en bénéficiant d'un service amélioré dans son ensemble. Sur le plan des performances, l'environnement gère des charges de travail bien plus variées que ne le peut l'infrastructure existante. Le service informatique fonctionne mieux en soi, et se concentre davantage sur l'entreprise que sur la technologie.

Protection des données

Bien que ce ne soit pas la tâche la plus agréable au monde, il est primordial de protéger les données. Il est triste de constater que de nombreuses organisations ne font que le plus strict minimum pour protéger leurs données critiques. Et ce pour deux raisons : Une protection complète des données peut s'avérer vraiment très coûteuse et réellement complexe.

Pour assurer la protection des données dans un système existant, il faut prendre beaucoup de décisions et acheter une grande sélection de produits. En revanche, dans un environnement hyperconvergé, la sauvegarde, la restauration et la reprise après sinistre sont intégrées. Elles font partie de l'infrastructure et ne sont pas des ajouts après-coup de tiers qu'il faut intégrer.

Les avantages de l'hyperconvergence sont clairs :

- ✓ Sauvegarde et restauration complètes et reprise après sinistre d'un coût abordable.
- ✓ Protection efficace sans réhydratation et re-duplication des données, et l'utilisation inefficace des ressources qui s'ensuit.
- ✓ Une console centralisée unique qui permet au service informatique de réagir rapidement.

Chapitre 8

Sept façons d'appliquer l'hyperconvergence

Dans ce chapitre

- ▶ Consolidation et modernisation
- ▶ Préparation contre le pire

Le plus intéressant avec l'hyperconvergence, c'est qu'il ne faut pas nécessairement que vous remplaciez l'infrastructure existante pour qu'elle ait immédiatement de la valeur. Voici sept façons dont vous pouvez dès aujourd'hui tirer les avantages d'une infrastructure hyperconvergée :

- ✔ **Consolidation des serveurs et du datacenter** Vous vous lancez dans un nouveau projet de consolidation ou bâtissez un nouveau datacenter ? Les principaux fournisseurs d'hyperconvergence proposent des produits qui s'intègrent facilement avec votre environnement existant. La bonne solution hyperconvergée peut résoudre vos problèmes immédiats et apporter d'importants avantages.
- ✔ **Modernisation technologique harmonieuse** Le fait que son implémentation ne soit pas perturbatrice fait la beauté de l'hyperconvergence. L'environnement hyperconvergé fait partie de votre environnement global, vous pouvez donc introduire une nouvelle architecture tout en éliminant progressivement l'ancienne, la mise en œuvre et l'expansion se faisant en fonction des fonds disponibles. Si des applications de l'environnement existant ont besoin des performances de stockage de l'environnement hyperconvergé, rien ne les empêche de recourir à ces ressources.
- ✔ **Déploiement de nouvelles applications de 1^{er} niveau** Votre environnement existant convient-il pour de nouvelles charges de travail de 1^{er} niveau ? Au lieu de simplement rajouter des ressources sur un environnement obsolète, déployez la nouvelle charge de travail dans un environnement hyperconvergé pour tirer profit des avantages opérationnels inhérents. Avec le temps, vous pourrez commencer à amener le reste de votre infrastructure dans la même

architecture, avec autant d'efficacité que s'il s'agissait de briques LÉGO qui s'ajoutent facilement aux autres.

- ✔ **Déploiement d'une VDI** Des îlots de ressources sont en grande partie établis pour répondre aux besoins de l'infrastructure de bureau virtuel (VDI). Toutefois, la façon dont le service informatique implémente ces îlots de ressources signifie qu'ils sont à jamais séparés. En déployant votre projet de VDI dans une infrastructure hyperconvergée, vous ne rencontrez pas de problèmes de ressources qui vous forcent à créer ces îlots. Quand le projet de VDI est terminé et que le reste de votre environnement peut être renouvelé, vous pouvez tout faire facilement glisser vers l'environnement hyperconvergé. Pour en savoir plus sur la VDI, voir le chapitre 2.
- ✔ **Gestion des sites à distance** Dans un environnement hyperconvergé, toute l'infrastructure est contrôlée par un seul système de gestion. Les ressources à distance peuvent être gérées comme si elles étaient locales. Pas besoin de personnel à distance pour réaliser des opérations manuelles, comme réaliser des sauvegardes, créer des LUN (logical unit numbers) ou des politiques de qualité de service. La technologie de gestion efficace des données permet la simplification des sauvegardes dans le bureau à distance et l'envoi automatique de copies hors-site jusqu'au siège, un autre bureau à distance ou même le cloud. Il est ainsi possible de centraliser les ressources administratives et de fournir des économies d'échelle de personnel.
- ✔ **Réalisation de tests et développement** Un grand nombre d'organisations exploitent des environnements de test et développement (test/dév), pour éviter d'introduire un code erroné dans la production. L'hyperconvergence prend en charge les besoins de test/dév et de production, grâce à des outils de gestion qui peuvent vous aider à créer des séparations logiques entre ces fonctions.

Malheureusement, beaucoup d'organisations n'accordent pas d'importance au test/dév et l'exécutent sur un matériel de classe inférieure, ce qui n'est pas vraiment logique. Cette plus grande agilité informatique permettra le maintien en interne des développeurs au lieu qu'ils bâtissent leur propre shadow IT dans le cloud public.
- ✔ **Modernisation de la sauvegarde et implémentation de la reprise après sinistre** Si vous faites mal les sauvegardes ou la reprise après sinistre, courez (ne marchez pas !) vers l'hyperconvergence pour en faire votre architecture d'infrastructure. L'infrastructure hyperconvergée élimine la complexité qui peut être inhérente à ces opérations. En informatique, il n'y a qu'un mantra : la simplicité. Et l'hyperconvergence est un des moyens les plus simples de parvenir à vos objectifs en matière de sauvegarde et de reprise après sinistre.

HPE SimpliVity : La nouvelle puissance en infrastructure hyperconvergée

HPE offre aux entreprises la plate-forme hyperconvergée la plus puissante de l'industrie, réunissant les meilleurs services de gestion de données de leur catégorie et le serveur le plus vendu au monde.

L'infrastructure hyperconvergée de HPE SimpliVity livre les performances de classe professionnelle, la résilience et la protection de données dont les dirigeants de systèmes d'information ont besoin aujourd'hui, ainsi que l'agilité et le modèle économique de cloud que les entreprises exigent. L'infrastructure hyperconvergée de HPE est puissante, simple et efficace, caractérisée par une approche permettant l'achat de constituants de l'infrastructure informatique au rythme de la croissance de l'entreprise, une gestion axée sur les charges de travail, l'optimisation hyper-efficace de la capacité et des performances, avec intégration de la protection des données et de la résilience.

L'infrastructure hyperconvergée de HPE SimpliVity simplifie énormément l'informatique en livrant toute l'infrastructure et tous les services de données avancés pour les charges de travail virtualisées en un seul constituant hyperconvergé évolutif, disponible sur l'excellente plate-forme de serveurs HPE ProLiant. Elle économise aux organisations jusqu'à 55 % par rapport aux solutions alternatives dans le cloud public, elle augmente de 81 % le temps que le service informatique peut passer sur de nouveaux projets et sur l'innovation et améliore de 70 % en moyenne les opérations de sauvegarde/restauration et de reprise après sinistre, tout en éliminant les produits tiers de sauvegarde et de reproduction.

L'infrastructure hyperconvergente de HPE SimpliVity : plus d'agilité, d'efficacité et de résilience à moindre coût et complexité.

Pour de plus amples informations, visitez le site www.hpe.com/info/hc

L'hyperconvergence pour transformer votre datacenter, simplifier l'informatique et abaisser le coût total de possession

Vous avez convergé vos charges de travail, alors pourquoi pas votre infrastructure ? Regrouper l'ensemble de l'infrastructure en un seul système informatique hyperconvergé consommable facilite les opérations. L'hyperconvergence résout les problèmes d'interopérabilité, améliore la gestion, réduit les coûts et donne à votre entreprise la même efficacité que celle du cloud. Découvrez les tenants et les aboutissants.

- **L'hyperconvergence totale : comprenez les occasions perdues en vous limitant aux serveurs et au stockage**
- **Approche axée sur les machines virtuelles : tournez le dos à l'infrastructure à silos et à la gestion par le matériel**
- **Introduction d'un nouveau modèle économique : reproduisez le modèle basé sur la consommation du cloud dans votre datacenter**
- **Création de nouveaux niveaux d'efficacité : augmentez les performances, abaissez les besoins de capacité, améliorez l'agilité et réduisez considérablement les coûts**

Scott D. Lowe est l'ancien PDG, fondateur et directeur consultant du 1610 Group, et co-fondateur d'ActualTech Media. Scott a écrit des milliers d'articles pour diverses publications et il contribue régulièrement sur VirtualizationSoftware.com et EnterpriseStorageGuide.com.



Ouvrez ce livre et découvrez :

- Pourquoi la virtualisation sur une infrastructure à silos a engendré des problèmes inattendus
- Comment l'infrastructure hyperconvergente va bien au-delà de l'infrastructure convergée
- Comment le SDDC, l'ITaaS et le cloud impactent votre datacenter
- Les façons dont des composants informatiques préconfigurés permettent l'agilité, l'efficacité et des économies d'échelle et de coûts.

Allez sur Dummies.com pour voir des vidéos, des exemples pas à pas, des articles pratiques, ou pour faire des achats !

WILEY



Aussi disponible au format électronique

ISBN 978-1-119-47864-5
Ne peut être revendu

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.