

LIVRE BLANC

COMMANDITÉ PAR:



DATACORE

SÉRIE CONSACRÉE À UNE GESTION PLUS INTELLIGENTE DU STOCKAGE

par Trevor Pott

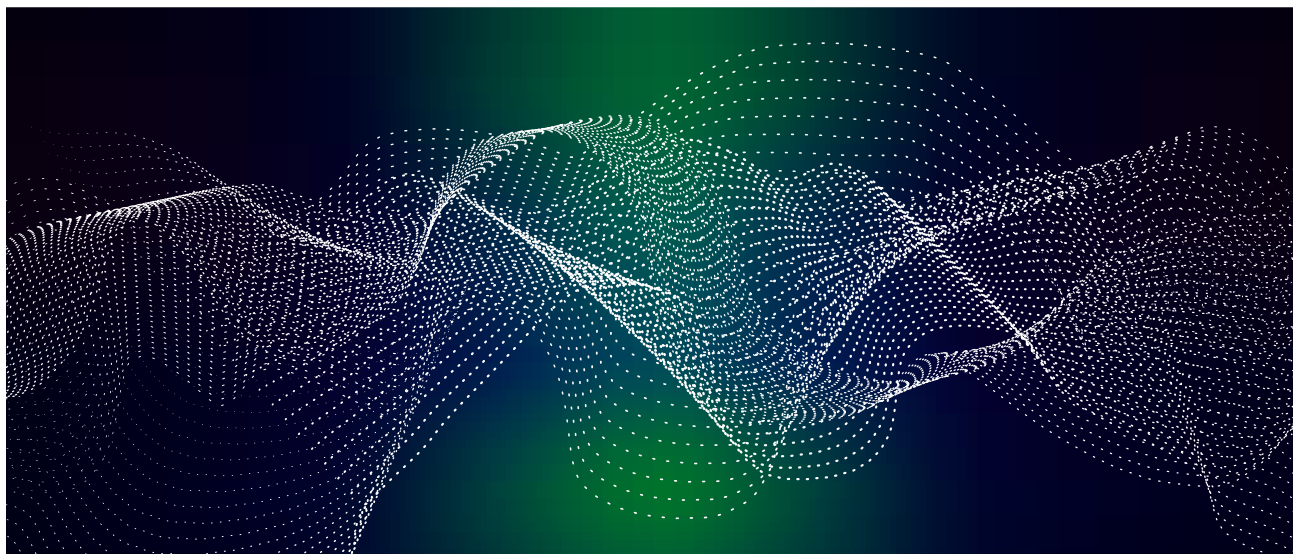
AU SOMMAIRE DE CE GUIDE :

- 2** Les fabrics de données et leur rôle dans la modernisation des matériels de stockage
- 6** La montée en puissance des infrastructures hyperconvergées
- 10** COMMENT MIEUX GÉRER LE STOCKAGE
Résoudre les 3 principales difficultés touchant vos données grâce au stockage défini par logiciel



LES FABRICS DE DONNÉES ET LEUR RÔLE DANS LA MODERNISATION DU MATÉRIEL DE STOCKAGE

Les fabrics de données permettent aux entreprises d'aller plus loin et de répartir leurs données sur de multiples dispositifs de façon à rendre redondantes les baies de stockage tout entières.



Comment moderniser le stockage ? Plus précisément, que se passe-t-il lorsque les contrats de support technique de matériels de stockage parfaitement fonctionnels arrivent à expiration ? Les entreprises doivent-elles prolonger les contrats de support à des prix éventuellement prohibitifs pour faire des économies d'investissement ou la génération de stockage précédente peut-elle rester utilisable sans support technique ?

Selon la personne à qui vous posez ces questions, vous obtiendrez une réponse différente. Les employés des fournisseurs vous donneront en général une série d'excellentes raisons pour ne jamais rien utiliser en

informatique sans contrat de support. De même, il y a apparemment de très bonnes raisons (fournies par qui ?) pour que les contrats de support des baies de stockage deviennent inéluctablement beaucoup plus chers dès la deuxième ou troisième année.

On vous fait comprendre, de façon pas toujours subtile, qu'il est de l'intérêt du client de renouveler son système de stockage tous les trois ans, même si celui-ci fonctionne parfaitement bien et répond aux besoins de l'entreprise. Bien que ce type de démarche puisse trouver un écho favorable auprès des entreprises où l'argent n'est pas un problème, jeter aux orties un système de stockage parfaitement

réparable pour dépenser des centaines de milliers, voire des millions de dollars pour acheter du matériel de rechange semble faire grincer des dents la plupart des autres.

Dépendance en matière de disques

Quel que soit le type de stockage utilisé, celui que l'on rencontre dans les centres de données comporte toujours deux types de composants. Il y a la partie qui stocke les données (les disques), qui peuvent être soit des supports magnétiques rotatifs (disques métalliques), soit des lecteurs à carte électronique (solid-state drives ou SSD). L'autre partie est celle qui choisit les disques chargés d'accueillir ces données. On l'appelle en général le contrôleur.

Les disques d'une baie de stockage d'entreprise ne sont en général pas très différents de ceux que l'on trouve dans les systèmes de stockage direct (NAS) grand public. Les disques d'entreprises se voient habituellement appliquer un microprogramme personnalisé, sans que l'on sache pourtant, après plusieurs décennies, si cela a une incidence ou non sur leurs performances ou leur fiabilité.

L'effet de ce microprogramme personnalisé est de donner aux fournisseurs de stockage un moyen d'empêcher les entreprises de remplacer tout simplement les disques de leurs baies de stockage par des modèles génériques du commerce. Un grand nombre de baies de stockage (surtout les plus coûteuses) refuseront les disques dépourvus de ce microprogramme, même si ceux-ci ont été achetés sous la mention « disque pour entreprise » ou « pour datacenter ».

Avec des dispositifs de stockage grand public, si les disques commencent à donner des signes de faiblesse ou si l'espace disponible commence à manquer, il suffit de se connecter sur Amazon pour acheter une douzaine de disques de rechange. Le fournisseur d'une solution de stockage grand public gagne sa vie en vendant le contrôleur et utilise des disques durs grand public non modifiés. Qu'est-ce qui empêche donc les fournisseurs de stockage d'entreprise d'en faire autant ?

Les raisons de remplacer le matériel de stockage

Les raisons invoquées par les fournisseurs de stockage d'entreprise en matière de remplacement des disques sont complexes. Les vendeurs de stockage professionnel ne peuvent pas se permettre de vendre des baies dont la solidité ne serait pas à toute épreuve. Le grand public et

les petites entreprises peuvent tolérer des taux de pannes et d'interruptions de service qui seront inacceptables pour les grandes sociétés. Cela signifie, entre autres, que les fournisseurs ne veulent pas que quelqu'un utilise des disques dont la compatibilité avec le matériel et le logiciel du contrôleur n'a pas été testée à 100 %.

Le fait que les disques grand public ne soient pas homogènes complique encore les choses. Un disque Western Digital Gold, par exemple, est considéré par Western Digital comme étant de classe entreprise. Si vous recherchez un disque dur haut de gamme à installer dans un serveur ou à placer dans votre NAS grand public, c'est le disque qu'il vous faut. Mais Gold est une marque, pas une gamme homogène de disques identiques.

Même dans le cadre d'une capacité spécifique, par exemple 12 To, ces disques sont produits en un grand nombre de séries, chacune présentant ses particularités et faiblesses propres. Western Digital modifie par ailleurs constamment la conception et le microprogramme de ces disques. Le résultat est que sur l'ensemble des disques Western Digital Gold de 12 To, il peut exister des douzaines de modèles différents, tous pourtant estampillés de la même marque.

Si le stockage grand public peut s'en accommoder, pourquoi les entreprises ne le pourraient-elles pas ? La réponse est que le stockage d'entreprise peut tout à fait le supporter, mais que les fournisseurs de stockage ne veulent pas avoir à gérer le chaos susceptible d'en résulter. Ils ont pour cela une bonne raison : les taux de pannes peuvent parfois atteindre des niveaux invraisemblables. Parmi les principaux modèles présentant un taux élevé de pannes, trois viennent immédiatement à l'esprit. Ce sont, dans l'ordre chronologique inverse, le [Seagate ST3000DM001](#) (surnommé ironiquement le « Failcuda », le catastrophique OCZ et [l'IBM Deskstar 75GXP](#) (surnommé le « Deathstar » ou « Étoile de la mort »). Le désastre qu'est l'OCZ est notable parce qu'il nous montre que les SSD peuvent présenter autant de problèmes que les disques rotatifs.

Des rumeurs circulent à propos d'autres modèles de disques dont les taux de pannes leur vaudraient de figurer sur cette liste, mais il est difficile d'obtenir des informations sur les taux de pannes des disques. Les fournisseurs de stockage, à l'exception notable de Backblaze, gardent jalousement le secret sur ces données et il faut en général qu'un juge en ordonne la divulgation pour qu'elles deviennent accessibles.

Devant le silence des professionnels, il est quasiment impossible de savoir quels modèles de disques sont susceptibles de poser des problèmes. Quel fournisseur voudrait en effet se trouver mis en cause par la fiabilité d'une baie de stockage exécutant des charges de travail stratégiques si les clients pouvaient tout simplement se mettre à y incorporer des disques à la fiabilité douteuse ?

Outre les préoccupations soulevées par les taux de panne des disques, les fournisseurs s'inquiètent également pour les contrôleurs. Le matériel physique présent dans la plupart des contrôleurs n'est en aucun cas différent de celui qui se trouve dans un serveur x86 moyen. Autrement dit, la plupart de ces dispositifs dureront au moins 10 ans. Or, un contrôleur conçu pour fonctionner avec des disques de 1 à 6 To risque de montrer des faiblesses si vous commencez à bourrer remplir votre baie avec des disques de 12 To.

Les baies de stockage possèdent une RAM qu'elles utilisent pour la mise en cache et pour exécuter des fonctions de contrôleur de stockage telles que la déduplication, la compression, le chiffrement, etc. Dans les baies de stockage d'entreprise, une partie, voire la totalité de cette RAM peut même être composée de RAM non volatile(NVDIMM). Celle-ci a été conçue pour que les contrôleurs puissent utiliser le cache en écriture sans avoir à se soucier des pertes de données dues aux pannes d'alimentation.

Avec l'installation de disques plus volumineux que ceux pour laquelle la baie a été conçue, le contrôleur risque de ne plus disposer d'assez de RAM pour faire correctement son travail. Dans le cas de contrôleurs équipés de NVDIMM, les cycles de remplacement triennaux s'expliquent en partie par le fait que les batteries et la mémoire flash qui permettent aux NVDIMM d'être non volatiles ont une durée de vie beaucoup plus brève que celle du reste du contrôleur.

Fabrics de données

Du point de vue du client, la dépendance vis-à-vis du fournisseur de disques est agaçante. On dirait qu'il ne s'agit de rien d'autre qu'une tentative cynique des fournisseurs d'extorquer de l'argent à un public captif et de traiter leur base de clientèle comme une vache à lait. Aux alentours de 2009, un grand nombre de start-up se sont mises à émerger en proposant de nouvelles approches du stockage.

La plupart d'entre elles cherchaient à exploiter le mécontentement de clients irrités par ce statu quo.

Pour compliquer encore une histoire qui ne se résume donc pas à une simple lutte de David contre Goliath, les fournisseurs de stockage avaient réellement d'excellentes raisons pour contraindre les clients à subir le verrouillage des disques, ainsi que d'autres pratiques visant à imposer des cycles de remplacement rapides. En tout cas, ces raisons étaient excellentes d'un certain point de vue.

Les fournisseurs de stockage vendent leurs produits en supposant qu'ils constitueront l'unique solution de stockage utilisée pour une charge de travail donnée.

Les clients sont censés exécuter des sauvegardes, mais la baie de stockage elle-même doit pouvoir fonctionner en permanence et offrir un stockage d'une robustesse irréprochable. Des vies peuvent en effet en dépendre. Cette approche du stockage a été établie avant l'existence des fabrics de données, et des générations entières d'ingénieurs ou d'administrateurs et de responsables du stockage ont été formées à ce type de raisonnement.

Or, les fabrics de données sont maintenant là. Une baie de stockage dispose de lecteurs redondants pour se protéger contre les pannes de disques, et même de contrôleurs redondants pour se prémunir contre les pannes de contrôleurs. Les fabrics de données permettent aux entreprises d'aller plus loin et de répartir leurs données sur de multiples dispositifs de façon à rendre redondantes les baies de stockage tout entières.

La plupart des fabrics de données offrent bien plus de fonctionnalités que la simple possibilité d'écrire sur plusieurs dispositifs de stockage afin de protéger les données contre la panne d'une baie entière. Les fabrics de données réévaluent en permanence les performances de tout le stockage du fabric, puis placent les données sur le système de stockage approprié en fonction des paramètres de profil des charges de travail utilisant ce stockage.

Si une baie affiche des performances insuffisantes (si, par exemple, vous l'avez bourrée rempli de disques d'une capacité supérieure à celle pour laquelle elle était conçue), le fabric la traite comme une destination de stockage à froid.

Il n'y placera donc que des blocs de données dont il a été déterminé qu'ils seront vraisemblablement peu consultés ou qui font partie d'une charge de travail dont le profil de stockage indique qu'elle est insensible à la latence.

Frankenstockage

Tout l'intérêt des fabricants de données est que vous pouvez utiliser n'importe quel système de stockage. Souhaitez-vous installer un stockage bon marché utilisant la gamme Supermicro [au prix ridiculement bas](#) ? Allez-y : vous trouverez des modèles qui vous permettront d'entasser [90 disques durs de 3,5 pouces](#) dans un rack de 4 u. Ou que diriez-vous de [48 disques NVMe](#) dans un serveur 2 u ?

Supermicro est un exemple facile de solutions bon marché imbattables, mais la plupart des fournisseurs de serveurs offrent [des solutions Opencompute](#) et le stockage Opencompute est désormais une réalité. On trouve aussi le [Backblaze Storage](#) Pod pour les utilisateurs qui recherchent un rapport qualité/prix maximum pour leurs modèles de stockage bon marché.

Le présupposé fondamental des fabricants de données est que vous assemblez une série de disques magnétiques à des fins de stockage en masse, puis que vous ajoutez des SSD (et d'ailleurs, de plus en plus, des SSD NVMe) pour un stockage performant.

Si vous avez besoin de davantage de capacité, il vous suffit d'ajouter des disques magnétiques. Si vous avez besoin de performances supérieures, vous ajoutez des SSD. Vous laissez le fabricant décider de l'endroit où placer les données et, en ajoutant simplement ce dont vous avez besoin, vous répondez à la question.

Mais il n'est pas nécessaire que les fabricants utilisent du stockage bon marché. Un grand nombre de fabricants de données autorisent les administrateurs à ajouter tous les types de stockage qui leur tombent sous la main. Vous avez une ancienne baie de stockage dont le support a expiré ? Ajoutez-la au fabricant, chinez des disques de rechange sur eBay et poussez cette baie jusqu'à ses limites. Vous stockez sur le

cloud ? Saisissez vos identifiants et hop : ça y est, vous avez de la capacité supplémentaire.

Si et quand votre baie hors support finit par vous lâcher, pas de problème : les données qu'elle hébergeait sont répliquées ailleurs dans le fabricant, lequel continuera à servir les charges de travail sans interruption. La plupart des fabricants sont capables de détecter la perte de la baie et d'entreprendre la création de copies supplémentaires des données qu'elle contient pour compenser ce stockage défaillant.

Les administrateurs de stockage classique désignent souvent les fabricants sous le terme moqueur de « Frankenstockage ». Les fournisseurs de fabricants sont terrifiés par le terme parce qu'il leur paraît

péjoratif, mais en réalité, l'analogie est très juste. Les fabricants de données permettent aux entreprises de piocher autant de solutions de stockage différentes qu'elles le souhaitent auprès d'un nombre illimité de fournisseurs et de les assembler en un résultat nettement supérieur à la somme de ses parties. Tout comme le monstre de Frankenstein, les fabricants de données restent profondément incompris et suscitent craintes, incertitudes et doutes.

Toute technologie informatique nouvelle rencontre des résistances, mais les fabricants de données ont aujourd'hui dépassé le stade de projet scientifique. Le marché compte de nombreux fournisseurs qui offrent des solutions éprouvées et qui commencent à avoir un impact concret sur la façon dont les entreprises gèrent le renouvellement de leurs systèmes de stockage. Si vous envisagez une nième refonte en profondeur de vos systèmes, c'est une technologie qu'il vous faudra prendre en compte.

LA MONTÉE EN PUISSANCE DES INFRASTRUCTURES HYPERCONVERGÉES

Implications pour votre plan de stockage de données.

Un grand nombre des fournisseurs réputés

d'infrastructures hyperconvergées (hyper-converged infrastructure ou HCI) ont été fondés en 2009. Nous avons donc 10 ans de recul pour évaluer cette technologie. Chacun sait qu'il s'agit de stockage et de calcul réunis dans un même système, mais qu'en est-il exactement ? Comment la HCI s'intègre-t-elle dans un environnement désordonné, mal optimisé et déjà occupé. Se montre-t-elle à la hauteur de la rumeur flatteuse dont elle fait l'objet ?

Pour comprendre la HCI, il faut connaître un peu l'évolution du stockage, ou tout au moins les formules ronflantes qu'elle a suscitées ces 10 dernières années. Lorsqu'elle est apparue en 2009, le terme que tout le monde

avait à la bouche à l'époque était « stockage défini par logiciel » (software-defined storage ou SDS).

Le SDS est l'archétype du mot à la mode puisqu'il est à peu près aussi dénué de sens que de parler de « structure bâtie par l'homme ». Certes, l'expression de structure bâtie par l'homme rétrécit quelque peu le champ des recherches. On comprend bien qu'il ne s'agit pas du nid sophistiqué construit par l'oiseau charpentier, ni d'une termitière, mais « structure bâtie par l'homme » reste une définition trop vaste pour être d'une quelconque utilité concrète.

Aujourd'hui, la quasi-totalité du stockage est définie par logiciel. C'est d'ailleurs aussi le cas d'une bonne partie du stockage de données physique : j'ai moi-même mis en place



des solutions d'archivage de documents papier qui utilisent des codes-barres, des codes QR ou le NFC pour indiquer les lieux de stockage et j'ai même contribué à automatiser la gestion du cycle de vie de ces documents.

Ce que le SDS devait signifier à l'origine était « des logiciels de stockage qui banalisent les fournisseurs de stockage de la même manière que VMware a banalisé les fournisseurs de serveurs ». Au bout du compte, l'expression a fini par désigner « tout type de stockage autre qu'une baie de stockage classique », ce qui ne veut rien dire.

Une baie de stockage de 2018 n'a plus grand-chose à voir avec son équivalent de 2009, mais la plupart des informaticiens refuseront pourtant de la qualifier de SDS en raison de sa filiation. Inversement, une solution HCI livrée par un fournisseur qui n'autorise ses clients à utiliser qu'un jeu étroit de nœuds que lui-même fournit sera malgré tout largement considérée comme du SDS.

Pour être encore plus clair : SDS est un terme en vogue qui, en réalité, a été forgé par des start-up pour dire plus poliment : « EMC, NetApp, IBM, HP et les autres sont trop chers, vous feriez mieux d'acheter vos solutions de stockage chez nous ». Il n'a aucune autre signification concrète.

Fabrics de données

Le concept que le SDS était censé défendre, c'est-à-dire la banalisation du stockage, reste pertinent. Pendant un temps, l'expression « virtualisation du stockage » a été utilisée. L'intention était de dresser un parallèle avec la banalisation des fournisseurs de serveurs par VMware, avec l'idée que le stockage pourrait être déplacé d'une solution à une autre aussi facilement que l'on pouvait déplacer des machines virtuelles entre les hôtes d'un cluster.

Cela n'a jamais pris, en grande partie parce que VMware était meilleur en virtualisation réelle du stockage que la plupart de ces start-up. vSphere possède une fonctionnalité appelée « storage vMotion » qui permet de déplacer facilement votre stockage entre diverses solutions pour peu que vos charges de travail soient virtualisées avec VMware. Il existe également les volumes virtuels, ou VVOL, qui sont censés faciliter la gestion des LUN et dont le succès (ou l'échec) est un sujet que nous réserverons plutôt à un autre débat.

Le concept d'origine du SDS est réapparu sous la forme du fabric de données. Un fabric de données est une application distribuée qui rassemble en un pool unique tous les éléments

de stockage qui lui ont été consacrés. Le logiciel répartit ensuite le stockage en fonction des paramètres qui lui ont été attribués et le met à disposition des utilisateurs.

Avec des fabrics de données ouverts, tous les types de stockage peuvent être utilisés : sur le cloud, bon marché, sur des serveurs locaux, des réseaux de zones de stockage locales (SAN), du stockage en réseau (NAS), que sais-je encore. Ce stockage peut alors être mis à disposition de la manière que le fournisseur de fabric de données choisira d'autoriser : iSCSI, Fibre Channel, SMB, NFS, Object ou même apparaissant aux nœuds comme du stockage local direct. Un fabric de données réellement ouvert peut utiliser tous les types de stockage, n'importe où, et livrer les données à la consommation de toutes les manières souhaitées.

Bien que la plupart des fabrics de données aient la capacité d'être aussi ouverts que ce que nous venons de décrire, peu le sont en réalité. Rien n'empêche un fournisseur de fabric de données d'intégrer n'importe quelle source de stockage ou d'émettre le stockage sous le format de son choix.

La difficulté pour concevoir un logiciel de fabric de données est de créer les éléments qui décident où placer les différents blocs de données et d'appliquer des fonctionnalités de stockage d'entreprise telles que la compression, la déduplication, la prise d'instantanés, le clonage rapide, le chiffrement, etc. À côté de cette difficulté, consommer un nouveau type de stockage ou ajouter un shim d'émission qui permet de le consommer sous un format différent n'a rien de compliqué.

Les fournisseurs, toutefois, restreignent les capacités de leurs fabrics de données pour les mêmes raisons que celles qui incitent les fournisseurs de stockage classique à pousser leur clients à dépendre de leurs disques et à encourager [le renouvellement du matériel tous les trois ans](#). Ce qui nous amène à la HCI.

HCI-Plus

Le fournisseur de votre logiciel de fabric de données est important car c'est lui qui détermine la flexibilité de ce fabric. Prenons la HCI. Il s'agit essentiellement d'un fabric de données composé de serveurs qui disposent d'un stockage local dont ils font don au fabric et qui utilisent leur capacité de calcul restante pour exécuter des charges de travail.

Bien qu'il n'y ait pas de raison interdisant à la plupart des solutions HCI d'intégrer d'autres sources de stockage dans

leur fabric de données, nombre d'entre elles sont pourtant restrictives. Les clients ne sont autorisés à ajouter aux cluster qu'une classe réduite de nœuds, la taille des clusters est confinée à un petit nombre de nœuds et les nœuds eux-mêmes sont rarement personnalisables en profondeur.

Les fabrics de données plus ouverts autorisent une approche dénommée « HCI-Plus ». Avec ces fabrics de données ouverts, les clients peuvent acheter des serveurs équipés de stockage et créer des clusters qui se comportent exactement comme une HCI. Ils peuvent également ajouter à l'ensemble leurs anciennes baies de stockage ou leur stockage sur cloud et les utiliser pour du stockage à froid, d'archivage, d'instantanés, etc.

Une utilisation concrète des fabrics de données

On peut raisonnablement se demander si une approche HCI+ du fabrics de données présente même un quelconque intérêt. Ça a l'air formidable sur le papier, mais si vous devez construire un fabrics de données, pourquoi ne pas tout simplement recourir exclusivement à des serveurs bons marché et utiliser une approche HCI classique ? Peut-être aussi utiliser du stockage sur cloud comme destination hors site pour les instantanés ou le stockage à froid, mais ajouter autre chose a-t-il réellement un sens ?

La réponse à cette question est compliquée. Si vous concevez un déploiement ex nihilo, une approche HCI+ a bien peu d'intérêt. La HCI marche très bien, au point que même des capacités basiques de réplication ou de sauvegarde sur le cloud attirent assez peu ses utilisateurs.

Les déploiements ex nihilo, il faut bien le dire, ne constituent pas la majorité des environnements. La plupart des entreprises possèdent un mix informatique assez désordonné. Différents niveaux ou différentes classes d'infrastructures appartenant à différents services sont tous remplacés à des moments différents. À la suite de fusions et acquisitions, les entreprises gèrent souvent plusieurs datacenters, chacun conçu de façon entièrement distincte et présentant une approche originale de l'informatique.

De nombreuses entreprises ne sont pas encore prêtes à adopter la HCI (et encore moins les fabrics de données) pour leurs charges de travail stratégiques. Pour elles, 10 ans ne représentent pas un recul suffisant pour qu'une technologie réponde à leurs exigences de

fiabilité et elles préféreront en rester à des solutions éprouvées, fiables et chères pour les prochaines années.

Pour autant, même dans les entreprises les plus traditionalistes, on constate en général une volonté de rechercher l'efficacité partout où elle existe. Les charges de travail non stratégiques, de développement ou de test, ainsi que le stockage des archives, représentent autant de domaines dans lesquels les administrateurs sont plus enclins à essayer des technologies « nouvelles ». C'est là que les fabrics de données (et plus spécifiquement l'approche HCI+) rencontrent le plus d'intérêt.

Le problème des HCI classiques

Le problème avec les solutions HCI classiques est qu'elles sont très restrictives en matière de composition des clusters. Le rapport entre la capacité de stockage, le nombre de cœurs de processeurs et la RAM est tout sauf flexible. En outre, les nœuds individuels qui composent le cluster ne peuvent avoir qu'un nombre restreint de SSD (et de SSD NVMe), ce qui limite les performances de stockage sur un nœud donné.

Tant que vous savez exactement ce que vous voulez placer sur un cluster donné avant de l'acheter, tout va probablement bien. Vous pouvez spécifier la capacité et les performances de ce cluster de façon à répondre précisément à vos besoins. Dans la réalité, cependant, les priorités changent, de nouvelles charges de travail sont introduites et celles attribuées à la plupart des clusters au moment de leur retrait n'ont plus rien à voir avec les intentions d'origine.

Confrontés à ce type de scénario, les administrateurs réagissent en général en surdimensionnant leurs clusters. Le surdimensionnement est certes une tradition à laquelle il a été sacrifié de tout temps, mais il tend à contredire l'objet même de la HCI. L'argument en faveur de la HCI est en général qu'elle vous permet d'étendre votre cluster au fur et à mesure des besoins, ce qui représente une source d'économies.

Contrairement à des fabrics de données plus ouverts, les solutions HCI classiques ne vous permettent cependant pas d'ajouter tout simplement quelques nœuds NVMe pour améliorer les performances, ni un gros boîtier plein de disques magnétiques pour ajouter de la capacité. C'est là qu'intervient la solution HCI-plus.

Non seulement une approche de type HCI-plus vous permet de créer des clusters HCI aux nœuds diversifiés, mais elle permet aussi de récupérer toutes les anciennes baies de

stockage éliminées du niveau stratégique et de leur donner une deuxième vie en les ajoutant aux fabric de données du niveau non stratégique.

Le loto des expressions ronflantes

Qu'on le veuille ou non, la nomenclature compte. Peu de fournisseurs (voire aucun) qui se targuent de fournir des solutions HCI se risquent à adopter le HCI-plus. Ceux qui proposent un fabric de données plus ouvert que la HCI classique n'ont pas encore pris de décision collective pour dénommer ce qu'ils font, en partie parce qu'il y a place pour de larges variantes entre la HCI et un fabric de données totalement ouvert.

L'incapacité des fournisseurs à s'accorder sur une nomenclature a singulièrement compliqué la tâche consistant à expliquer aux clients ce que font les fournisseurs et en quoi ils diffèrent de leurs concurrents. Cette difficulté pose à son tour des problèmes aux clients pour trouver la solution qui leur correspond le mieux. Cela explique en partie pourquoi le marché du stockage est une telle pagaille.

Ce qu'il est important de se rappeler, c'est que, bien que de nombreux fournisseurs de HCI se présentent comme le nec plus ultra de la modernité, sous-entendant que les baies de stockage classiques sont une solution dépassée, la HCI elle-même a déjà près de 10 ans. Ce n'est plus une nouveauté et dans l'intervalle de ces 10 ans, de nombreux fournisseurs de HCI se sont révélés tout aussi prompts à enfermer leurs clients dans des solutions exclusives, à surfacturer leurs services et à restreindre la liberté de choix de leurs clients que les fournisseurs qu'ils cherchaient à remplacer.

Les fabric de données ouverts représentent la même promesse que cette révolution du stockage que l'on nous annonçait il y a 10 ans. Ils sont ce que la HCI aurait dû être. Reste encore à voir si les fournisseurs actuels de HCI ouvriront leurs produits de manière à offrir au moins une approche HCI-plus ou s'ils disparaîtront pour céder la place à une nouvelle génération de fournisseurs de fabric de données ouverts, impatients de définir leur propre vocabulaire et de laisser leur propre marque dans l'histoire de l'informatique.

COMMENT MIEUX GÉRER LE STOCKAGE

Résoudre les 3 principales difficultés touchant vos données grâce au stockage défini par logiciel

Le secteur informatique fait face à de réelles difficultés. Parmi les principales, on peut notamment citer la nécessité de maintenir un service ininterrompu, d'évoluer avec des demandes de capacité sans cesse croissantes, de déployer des architectures hautes performances et de réduire les coûts.

Le problème central : LES DONNÉES

Qu'il s'agisse de se conformer aux nouvelles réglementations, de déployer de nouvelles applications ou de faire face à une croissance rapide, vos données sont au cœur de toutes les fonctions liées aux opérations et aux services. Les données stratégiques doivent toujours être accessibles et immédiatement récupérables. Si ce n'est pas le cas, l'entreprise peut y perdre une part importante de son activité.

En outre, les stratégies actuelles de conservation des données informatiques font du stockage et de la gestion des tâches considérables. Les systèmes de stockage informatique doivent évoluer comme jamais auparavant. Et pour couronner le tout, les responsables informatiques doivent travailler avec des budgets particulièrement serrés.

La création d'une solution de stockage classique qui offre une accessibilité, une évolutivité et des performances de classe entreprise peut déjà être compliquée. Mais il faut également être en mesure de répondre en un clin d'œil aux fluctuations de ces mêmes exigences. Une telle réactivité nécessiterait un investissement financier et humain conséquent, sans même parler de l'entretien.

La solution DataCore : le stockage défini par logiciel

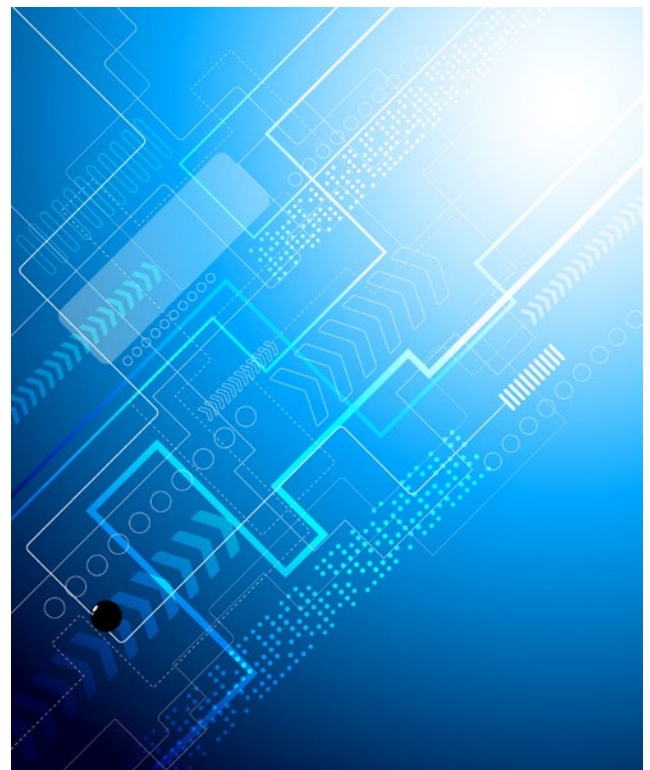
SANSymphony™, la solution de stockage défini par logiciel de DataCore™, est la plateforme de services de stockage d'entreprise la plus solide du marché. Elle offre un accès ininterrompu aux données, des capacités impressionnantes d'élargissement de l'architecture et un niveau de performances

applicatives qui change la donne. Le logiciel crée une couche de virtualisation transparente au-dessus de divers systèmes de stockage pour optimiser la disponibilité, l'évolutivité et les performances de toutes les ressources de stockage.

Étudions la façon dont SANSymphony résout l'ensemble de ces difficultés, simplifie la gestion du stockage à l'échelle de l'entreprise et vous rend la maîtrise de votre stockage.

Difficulté n° 1 : disponibilité des données

La disponibilité des données vise à garantir l'accessibilité des données, même en cas de panne catastrophique au sein de l'architecture de stockage due à des dysfonctionnements



matériels, des pannes de sites, des sinistres régionaux ou des erreurs des utilisateurs. SANsymphony assure la disponibilité continue des données grâce à des fonctions telles que la mise en miroir synchrone, la protection continue des données (CDP), les canaux de communication multichemins et la réplication des sites distants pour les opérations de reprise après sinistre. DataCore apporte également son aide pour la mise en place de l'infrastructure et des procédures de stockage afin que vous disposiez d'une solution complète.

Mise en miroir synchrone

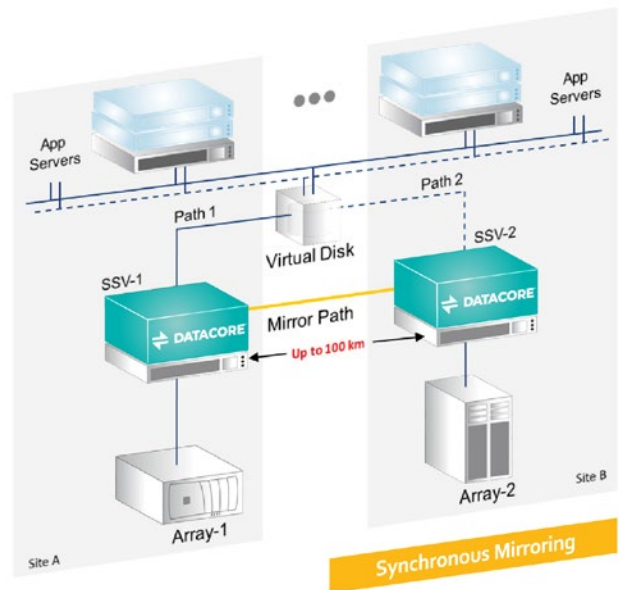
La nécessité de protéger les données stratégiques est extrêmement importante pour votre entreprise. Avant de voir comment vous pouvez y parvenir, penchons-nous sur la distinction entre tolérance aux pannes et haute disponibilité.

La tolérance aux pannes assure une protection contre les pannes matérielles au niveau des composants du système de stockage tels que vos disques durs ou vos éléments d'alimentation électrique. Tous les systèmes informatiques de production devraient bénéficier de capacités de tolérance aux pannes afin de se prémunir contre les pannes matérielles qui se produiront à coup sûr. Mais si les mesures de protection s'arrêtent à la tolérance aux pannes, alors le composant le plus important a été ignoré : vos **DONNÉES** !

SANsymphony protège contre les pannes catastrophiques des systèmes de stockage susceptibles de se produire lors de défaillances environnementales (par exemple, panne de courant, incendie, perte de la climatisation de la salle informatique, etc.). Les mesures de tolérance aux pannes sont incapables de protéger efficacement contre ce type de sinistre. La haute disponibilité ajoute la redondance des données à la redondance des composants matériels afin d'optimiser la disponibilité de vos données et de garantir le fonctionnement ininterrompu de vos applications.

La solution de stockage défini par logiciel SANsymphony assure la haute disponibilité grâce à la mise en miroir synchrone de chaque bloc de données sur un deuxième nœud SANsymphony entièrement actif. Ces nœuds entretiennent des copies redondantes de sorte que même si un pan entier de l'infrastructure de stockage devient inutilisable, les applications ne subiront aucune interruption et les données resteront toujours disponibles.

Pour déployer une solution hautement disponible, il y a quatre principes à respecter :



› **Redondance de bout en bout** : les systèmes de stockage sous-jacents doivent comporter un niveau raisonnable de tolérance aux pannes (redondance au niveau des composants matériels) associé à des capacités de haute disponibilité (redondance au niveau des données).

› **Autonomie des sous-systèmes** : les systèmes de stockage sous-jacents doivent travailler indépendamment, dans l'ignorance de leur existence mutuelle.

› **Séparation des sous-systèmes** : les systèmes sous-jacents peuvent être éloignés les uns des autres de 100 km au maximum pour éviter qu'un incident survenant sur un site pèse sur la prestation des services de stockage sur l'autre.

› **Asymétrie des sous-systèmes** : le fabricant et/ou le modèle des systèmes de stockage sous-jacents existant de chaque côté peuvent être différents afin qu'un bogue ou un problème logiciel/matériel survenant sur un site n'ait pas de conséquences sur l'autre.

Protection continue des données

Un autre aspect important à prendre en compte pour protéger les informations stratégiques est l'utilisation de mesures de protection continue des données (Continuous Data Protection, CDP). La CDP offre des capacités de retour inférieures à la seconde pour tout volume compatible CDP. Ce système offre le meilleur objectif possible de point de récupération (recovery point objective ou RPO) et de temps de récupération (recovery time objective ou RTO) en cas de perte de données involontaire due à un logiciel malveillant ou à une erreur d'utilisation.

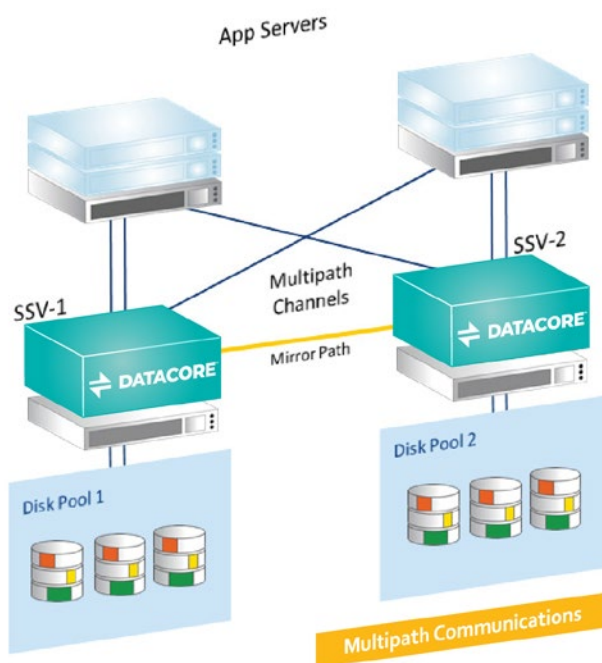
La CDP peut également servir à ramener des volumes à un point temporel antérieur si une sauvegarde a été manquée. Elle est d'une grande utilité pour les entreprises soumises à des règles strictes de conservation des données et de conformité aux obligations de récupération.

Canaux de communications multichemins

La disponibilité des données concerne non seulement la protection des données, grâce à la mise en miroir synchrone et à la CDP, mais également l'accès continu aux données, grâce à des canaux de communications multichemins.

Si un canal de communication devient inutilisable pour une raison ou une autre, il faut que les données restent techniquement accessibles. Or, la largeur de bande totale des canaux utilisés pour accéder aux données se trouve réduite. Face aux exigences élevées des charges de travail applicatives modernes, cette situation peut gravement nuire aux performances et à l'expérience utilisateur.

SANsymphony est une solution logicielle de virtualisation du stockage totalement abstraite et indépendante. Cela signifie que vous êtes totalement libre d'adapter et de modifier votre architecture de stockage pour conserver vos niveaux de services dans des conditions de fonctionnement normales ou anormales.



Réplication de site distant

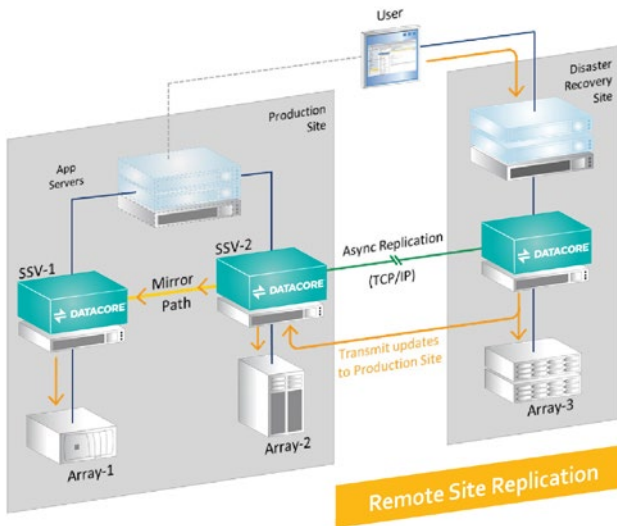
La dernière ligne de défense pour préserver la disponibilité des données est la réplication de sites distants au cas où une catastrophe naturelle toucherait une vaste région géographique. L'utilisation de sauvegardes pour restaurer tout un site n'est pas réaliste en raison du temps nécessaire à la récupération (en supposant d'ailleurs qu'il reste quelque chose à récupérer). La réplication asynchrone par SANsymphony, associée à la fonctionnalité de récupération avancée des sites (Advanced Site Recovery ou ASR), transforme un risque de cauchemar informatique en une procédure pouvant être gérée et testée. Ensemble, ces deux fonctionnalités peuvent faire toute la différence entre rester dans la course et faire faillite.

Dans le cadre de la réplication asynchrone, les données sont copiées sur le site distant au niveau des blocs et deviennent rapidement accessibles si le site de production devient indisponible. La réplication asynchrone offre le niveau d'intégrité le plus élevé et la solution de reprise après sinistre la plus flexible pour les systèmes d'exploitation et les applications et les données stratégiques.

Un plan de reprise après sinistre n'est pas complet sans la possibilité de tester la procédure de basculement et de garantir l'intégrité des données sur le site distant. Avec le stockage défini par logiciel de SANsymphony, vous pouvez tester le basculement de site à site aussi souvent que nécessaire, sans conséquences sur l'environnement de production et sans interrompre la procédure de réplication sur le site distant. Il suffit pour cela d'activer la fonctionnalité du mode test pour les volumes à vérifier. Une fois le test sur le site distant réalisé, désactivez le mode de test. SANsymphony ramène alors automatiquement la configuration du site distant dans le mode de fonctionnement normal.

Il existe de nombreuses méthodes pour répliquer les données sur un site distant, c'est la partie la plus facile. Mais que se passe-t-il si un basculement, puis un rebasculé deviennent nécessaires ? Voilà qui est beaucoup plus difficile.

ASR inverse automatiquement le sens de la réplication asynchrone lors d'une opération réelle de basculement ou de rebasculé afin que l'autre site reçoive toutes les mises à jour une fois les services restaurés. En outre, une resynchronisation totale de l'ensemble des données n'est pas nécessaire pour ramener le site d'origine en ligne. Cela évite une interruption de service inutile lors de la transition.



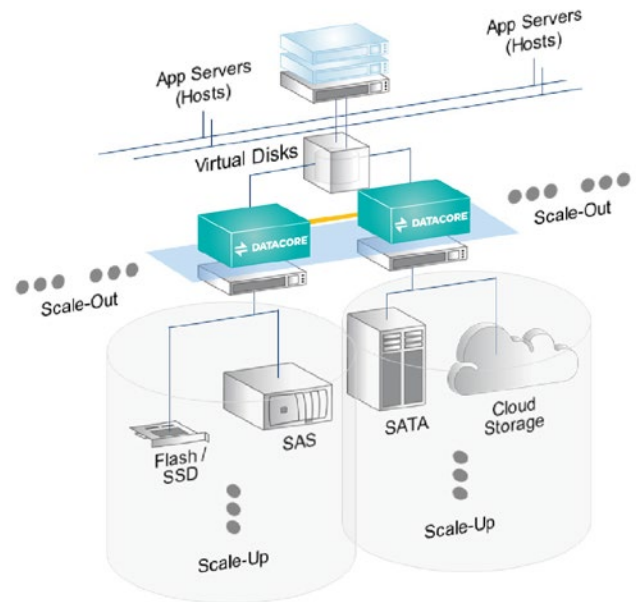
Difficulté n° 2 : utilisation et évolution de la capacité

SANsymphony veille à ce que vos données soient stockées efficacement en optimisant l'utilisation des ressources de stockage. Il y parvient grâce au thin-provisioning (ou provisionnement léger), qui n'alloue dynamiquement l'espace que lorsque les données sont écrites. Cela se traduit par une augmentation de l'utilisation du stockage et une récupération de la capacité (jusqu'à quatre fois). Le thin-provisioning supprime également les opérations de redimensionnement des volumes logiques puisque la présentation de volumes logiques de grande taille aux serveurs d'applications n'entraîne aucune pénalité. Au fil du temps, SANsymphony peut récupérer l'espace libéré par la suppression ou le retrait de certaines données.

La solution DataCore vous permet également d'adapter rapidement et en toute sécurité votre infrastructure de stockage, par scale up ou scale out, et de maîtriser l'évolution continue de la demande. Qu'il s'agisse d'ajouter des nœuds ou d'augmenter la capacité des nœuds existants, cela peut être fait rapidement et sans interruption de l'environnement de production.

Il y a différentes raisons de choisir de redimensionner les systèmes en ajoutant des serveurs (scale out) ou en remplaçant les serveurs existants par des modèles plus puissants (scale up). Selon le serveur matériel déployé pour servir de nœud SANsymphony, vous pouvez décider d'une évolution simple par scale up en ajoutant de la capacité de stockage (à la fois interne au serveur et externe), un cache haute vitesse, des canaux de communication, des processeurs, etc.

Si d'autres IOPS s'avèrent nécessaires et que le matériel existant atteint ses limites de capacité, d'autres nœuds peuvent être ajoutés pour exécuter les charges de travail applicatives. Quelle que soit la solution choisie, vous pourrez ajouter des ressources sans désactiver vos systèmes de production.



Difficulté n° 3 : les performances de stockage

Les systèmes informatiques de stockage doivent être capables de performances extrêmes pour suivre le rythme d'acquisition des données et le nombre croissant des applications.

SANsymphony offre des performances exceptionnelles grâce à deux méthodes principales : la mise en cache haute vitesse et la hiérarchisation automatique du stockage.

Mise en cache haute vitesse

La première méthode consiste à utiliser une RAM haute vitesse comme cache. Le logiciel SANsymphony utilise les processeurs, la mémoire et les ressources d'E/S du nœud pour exécuter des algorithmes de mise en cache multithread sophistiqués sur l'ensemble des dispositifs de stockage gérés. Le logiciel peut réserver jusqu'à 1 téraoctet (To) de RAM par nœud pour créer des caches de très grande taille à l'échelle du SAN.

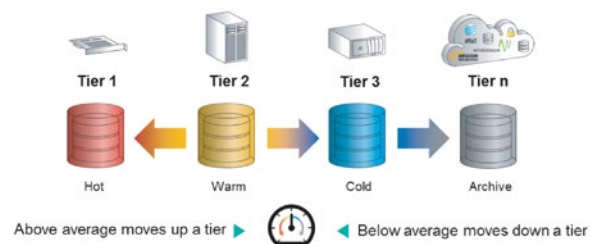
La mise en cache haute vitesse est essentielle pour conserver les performances applicatives puisque la RAM est considérablement plus rapide que les plus rapides technologies flash et qu'elle se trouve à proximité du processeur, ce qui

réduit la latence. C'est le composant de stockage le plus rapide de l'architecture, capable de multiplier par 10 les performances applicatives et de libérer les serveurs d'application, qui peuvent ainsi effectuer d'autres tâches. Il prolonge également la vie des composants de stockage magnétiques traditionnels en limitant l'usure résultant du disk thrashing.

Hiérarchisation automatique du stockage

La deuxième fonctionnalité favorisant l'accélération du stockage est la hiérarchisation automatique du stockage. Cette méthode vise à déplacer les données en temps réel vers des niveaux de stockage plus performants en fonction de la fréquence des accès. Associée à des dispositifs de stockage haute vitesse tels que la technologie flash, elle se traduit par une amélioration très sensible des performances applicatives.

La hiérarchisation automatique du stockage permet d'intégrer de façon transparente le stockage flash ou d'autres systèmes de stockage à haut débit dans l'architecture de stockage globale. Il suffit d'ajouter des composants de stockage flash au groupe de disques SANsymphony, aux côtés des disques magnétiques existants. SANsymphony déplacera automatiquement les données entre les diverses technologies de stockage au niveau des blocs des sous-LUN, offrant ainsi les meilleures performances aux applications qui en ont besoin.



Une excellente raison d'utiliser la hiérarchisation du stockage est qu'elle vous permet d'obtenir des performances identiques à celles d'une architecture entièrement flash sans avoir à installer cette architecture. En général, le stockage total comptera 8 à 10 % de dispositifs SSD haut débit. SANsymphony veillera à ce que les données chaudes (fréquemment consultées) résident sur le niveau le plus rapide, les données froides (consultées plus rarement) résidant sur les niveaux plus lents et moins coûteux.

Les statistiques du secteur montrent qu'en moyenne, seuls 15 à 20 % de l'ensemble des données d'une entreprise bénéficient du stockage flash puisque le pourcentage restant est considéré comme essentiellement dormant. Ainsi que le note un rapport d'analystes :

« Les pratiques d'accès, la valeur et les caractéristiques d'usage des données stockées dans des baies de stockage varient considérablement et dépendent des cycles d'activité de l'entreprise et de ses processus d'organisation du travail. En raison de cette large variabilité, les données stockées

sur des dispositifs de stockage ne peuvent pas l'être de façon économique et efficace sur les mêmes types, niveaux, formats ou supports de stockage. »

SANsymphony gère jusqu'à 15 niveaux différents de technologies de stockage, dont le stockage sur le cloud pour les données d'archivage à long terme rarement consultées. À mesure que des technologies plus avancées deviennent disponibles, les niveaux existants peuvent être modifiés en fonction des besoins et des niveaux supplémentaires intégrés afin de diversifier davantage l'architecture.

Cela se traduit au bout du compte par une amélioration des performances applicatives et par la baisse du TCO sur l'ensemble de l'architecture de stockage.

Conclusion

En informatique, chaque investissement doit entraîner un retour optimal. Les entreprises ont également un fort besoin de systèmes de stockage professionnels, fiables et performants, lesquels coûtent cher. SANsymphony vous permet de profiter du meilleur des deux mondes.

SANsymphony est une solution logicielle de virtualisation du stockage totalement abstraite et indépendante. Elle vous donne la maîtrise absolue de votre architecture de stockage, tout en vous offrant les plus puissantes fonctionnalités de stockage d'entreprise disponibles aujourd'hui.

Imaginez l'impact sur le ROI et le TCO d'une solution sur mesure qui vous permet d'utiliser la marque et le modèle de système de stockage de votre choix, tout en vous offrant des performances fulgurantes et les formes de protection les plus évoluées du marché pour sécuriser vos données. Avec en outre la possibilité, à mesure que les conditions évoluent, de développer sans interruption les ressources nécessaires pour répondre aux exigences de nouveaux environnements.

C'EST CETTE PUISSANCE QUE VOUS OFFRE DATACORE !

Trevor Pott est un spécialiste technologique réputé comptant plus de 20 ans d'expérience au service d'entreprises de tailles diverses en qualité d'informaticien, de conseiller, de consultant et d'auteur. Auteur technologique prolifique et consultant fournisseur très recherché, connaissant à la fois la technologie et ceux qui l'appliquent, Trevor est également administrateur système chez eGeek Consulting Ltd.

Consultez le site de DataCore dès aujourd'hui à l'adresse datacore.com.

