

STORAGE

LE MAGAZINE DU STOCKAGE INFORMATIQUE PROFESSIONNEL



LES STARTS-UPS FRANÇAISES DU STOCKAGE ET DE LA PROTECTION DES DONNÉES



JUIN 2018

VOL. 15

BLEU,
BLANC,
ROUGE...

LES OFFRES
DE STOCKAGE
EN CLOUD DES
ACTEURS FRANÇAIS

LES 8 START-UPS US
DU STOCKAGE
À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE
POUR HADOOP

S3 : UNE OPTION
DE STOCKAGE
DE PLUS EN PLUS
UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?



BLEU, BLANC, ROUGE...

PAR CHRISTOPHE BARDY

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

L'ADAGE « LA TECHNOLOGIE N'A PAS DE FRONTIÈRES » n'a jamais autant été remis en question qu'au cours des derniers mois. On a ainsi vu le gouvernement nippon surveiller d'un œil acéré le nom des acquéreurs de la division mémoire de Toshiba, le numéro deux mondial de la mémoire Flash, tandis que les États-Unis mettaient quasiment en faillite l'un des principaux équipementiers télécoms mondiaux, ZTE, en lui interdisant l'accès à toute technologie et composant électronique d'origine américaine pour avoir commis la boulette, en ces temps compliqués, de commercer avec l'Iran. Paradoxalement, ce n'est que par la grâce (présidentielle) de Donald Trump que ZTE se doit d'être encore vivant. Moyennant une copieuse amende, un remplacement de ses dirigeants et un engagement à acheter US, ZTE s'est vu accorder un répit alors qu'il avait littéralement été contraint de stopper ses usines faute d'accès à la technologie US. Gageons que le Chinois saura s'en souvenir et qu'il y a fort à parier que, d'ici à cinq ans, la part des composants US dans ses approvisionnements s'en trouvera radicalement modifiée.

Et les entreprises françaises me direz-vous ? Eh bien, elles restent largement dépendantes de technologies américaines pour une large part de leur IT et en particulier pour la composante infrastructure. Dans l'applicatif, quelques champions européens comme SAP, Sage ou Dassault Systèmes continuent à tenir la dragée haute à leurs homologues US, mais dans l'infrastructure, il est difficile de voir émerger des champions susceptibles de prendre le relais des fournisseurs américains.

DANS L'UNIVERS DU STOCKAGE, UN PETIT MAIS VAILLANT VILLAGE GAULOIS RÉSISTE AUX ROMAINS CALIFORNIENS.

Dans le stockage, les acteurs français ont souvent été précurseurs dans certains domaines de l'infrastructure. Pourtant, qui se souvient aujourd'hui de Seanodes, le pionnier français du stockage SAN en cluster, qui en 2002 a conçu l'équivalent de vSAN, la technologie de stockage distribué servant aujourd'hui de pilier à la stratégie hyperconvergente de VMware ? Faute de clients, faute d'investissements, la



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

société a coulé en 2009 alors qu'elle avait été l'une des pionnières du mouvement [software defined storage](#) et de l'[hyperconvergence](#).

Autre pionnier, encore vivant, malgré un parcours chaotique, Active Circle a été l'un des acteurs d'avant-garde du [stockage NAS](#) en cluster aux côtés de Spinnaker (racheté par NetApp) et bien avant les Exanet (racheté par Dell) et autres Ibrix ([racheté en 2009 par HPE](#)). La firme a toutefois peiné à vendre sa solution a raté son implantation aux USA et n'a dû son salut qu'à Oodrive qui l'a [rachetée](#) en 2014 et continue à développer sa technologie.

Nul n'est prophète en son pays, dit le dicton. Car pour une start-up française, convaincre les grands comptes français d'utiliser sa technologie a longtemps ressemblé à un parcours du combattant. Dans un pays notoirement averse au risque, les petits acteurs français, aussi innovants soient-ils, ont longtemps été snobés par les grands donneurs d'ordres, y compris publics, faute de ne pas s'appeler IBM, EMC ou HPE. Sans un marché national pour faire leurs premiers pas, les créateurs français de technologies de stockage ont souvent été contraints à l'exil (c'est le cas de DDN, dont les fondateurs français ont préféré s'implanter directement aux États-Unis) pour prospérer.

Les choses nous dit-on seraient en train de s'améliorer. Le besoin de réduire les coûts et d'innover, la quête de

technologies européennes pour se protéger des fourches caudines de l'oncle Sam et de ses velléités législatives transnationales (Patriot Act, Cloud Act) amèneraient certaines entreprises à regarder avec un nouveau regard des acteurs français et européens qu'elles n'auraient pas considérés il y a quelques années.

Dans ce numéro de Storage, nous avons donc tenté de faire un tour d'horizon des principaux fournisseurs français de solutions de stockage, histoire de donner un coup de pouce à des acteurs dont la technologie n'a souvent rien à envier à celle de leurs homologues internationaux. En parallèle, nous vous proposons un tour d'horizon de huit start-ups prometteuses dans le domaine du stockage. De façon anecdotique, ce panorama comprend un grand nombre de sociétés bien financées et nées en Israël, mais qui se sont projetées très tôt aux États-Unis et à l'international (ce qui confirme pour le coup que nul n'est prophète en son pays).

Enfin, nous vous proposons un point rapide sur l'évolution du stockage dans Hadoop et sur les travaux très actifs pour faire du [stockage objet](#), une alternative pratique au [stockage HDFS](#) dans les clusters Hadoop.

Bonne lecture estivale ■

CHRISTOPHE BARDY est rédacteur en chef adjoint du LeMagIT en charge des infrastructures informatiques.

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

LES START-UPS FRANÇAISES DU STOCKAGE ET DE LA PROTECTION DE DONNÉES

LeMagIT fait le point sur six sociétés françaises qui tentent de se faire une place sur le marché du stockage face aux géants mondiaux du secteur.

Par Christophe Bardy



LE MARCHÉ MONDIAL DU STOCKAGE et de la protection de données est actuellement dominé par des acteurs anglo-saxons. Mais une poignée d'éditeurs français proposent des alternatives séduisantes aux solutions élaborées par les fournisseurs américains. Si tous ces acteurs français entendent poursuivre leur croissance dans l'hexagone, ils affichent aussi pour la plupart des ambitions de développement international, en Europe et dans le monde.

Dans la plupart des cas, la technologie de ces sociétés n'a pas à rougir par rapport à celle de leurs concurrents internationaux. Sur le marché du stockage objet, Scalify fait partie des leaders mondiaux tandis qu'OpenIO est référencé par Gartner comme l'un des acteurs à suivre. Sur le marché du stockage NAS, Rozo Systems et Oodrive proposent des solutions innovantes dont les caractéristiques ont séduit

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

plusieurs grands clients en France comme à l'international. En matière de partage et de synchronisation de fichiers, les offres "on premises" (auto-hébergement) de Pydio et en mode [SaaS](#) d'Oodrive ont déjà conquis des milliers de clients désireux de trouver une alternative à Dropbox, Box, GSuite ou OneDrive. Enfin, dans le domaine du backup, le nouveau groupe Wooxo-Atempo propose un panel de solutions à même de répondre aux besoins de sauvegarde et d'archivage d'entreprises allant de la simple PME aux grands comptes de taille internationale.

À l'heure actuelle, tous les besoins de stockage d'une entreprise ne peuvent pas être satisfaits à 100 % par des solutions françaises, mais le périmètre fonctionnel couvert par les produits des éditeurs que nous avons recensés dans ce dossier est déjà large. Le mouvement vers le [software defined storage](#) pourrait de plus encourager l'émergence de solutions additionnelles dans l'avenir.

ON N'EST PAS FORCÉMENT PROPHÈTE EN SON PAYS

Paradoxalement, le caractère français de ces éditeurs ne leur a pas toujours simplifié la vie en France. Comme l'explique Luc d'Urso, le patron de Wooxo-Atempo, le fait d'être français n'est pas toujours un plus pour séduire des clients français. Il pointe également l'absence d'une réglementation de type Small Business Act à la française ou d'un Buy European Act qui encouragerait l'achat de

solutions nationales ou européennes.

Pydio a quant à lui choisi de gommer les attributs français de son offre, à tel point que la nouvelle version du site de l'éditeur a d'abord été développée en anglais. Ce choix est guidé par la volonté de se développer plus facilement à l'international. Le cofondateur de Pydio, David Gillard, note toutefois que le fait d'être français peut faciliter la vie sur le marché des administrations et qu'il arrive que les clients PME y soient sensibles. À l'inverse, il souligne qu'il faut parfois de gros efforts pour rassurer les grands comptes français sur la pérennité de sa solution lorsque l'on est une start-up française.

À l'inverse, Philippe Boyon, en charge du marketing produit d'OoDrive, souligne que c'est clairement un avantage d'être un fournisseur français pour les clients français. « Aujourd'hui, nous proposons des applications hébergées en France avec du stockage hébergé en France et c'est un critère de choix important pour nos clients. 80 % des acteurs du CAC40 sont ainsi nos clients et nous comptons énormément de clients parmi les administrations françaises. Pour certaines entreprises, le fait que nous soyons français et non soumis au [Cloud Act](#) est essentiel ».

■ OPENIO

Le français OpenIO est l'un des nouveaux venus dans le monde du stockage objet, mais sa technologie a des racines anciennes puisqu'elle a fait ses débuts il y a plus de



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

10 ans. [La technologie de stockage objet d'OpenIO](#) trouve son origine dans une solution de stockage logicielle développée sur mesure à partir de 2007 par Atos Worldline pour le stockage et l'archivage des services de messagerie électronique de Wanadoo (aujourd'hui Orange) puis également ceux de SFR.

En 2012, le code du logiciel est devenu Open Source à la demande d'Orange et en 2015, plusieurs des mainteneurs de ce code ont profité de la réorientation stratégique de Worldline vers les métiers du bancaire pour créer un fork et en faire un produit à part entière. En juin 2015, OpenIO était née. Depuis, la société n'a cessé de se développer en acquérant de nouveaux clients et en imaginant des services de données et des services de "compute" innovants autour de sa solution de stockage.

Un peu plus de deux ans après les débuts de la commercialisation de sa technologie, la firme compte plus de 25 clients dans le monde pour près de 30 Po de données stockées. Parmi eux figurent plusieurs références prestigieuses, dont SFR, Dailymotion ou le CEA. OpenIO, historiquement adossé au groupe français Okto — la maison mère de l'éditeur [Vade Retro](#), qui protège 265 millions de boîtes aux lettres mails dans le monde — a levé quelque 5 millions d'euros à la fin 2017 pour accélérer son développement.

L'éditeur a conçu une solution de stockage objet compatible avec les API Amazon S3 et OpenStack Swift et capable de délivrer des services de partage de fichiers NFS. Multi-plate-forme - x86 et ARM - et Open Source, le logiciel répond aussi bien aux besoins de développeurs, qu'à des besoins de stockages de données massifs en Pétaoctets. Selon les déploiements, la protection des données est assurée par [réplication](#) ou par un mécanisme d'[erasure coding](#) et le système intègre également des mécanismes sophistiqués de géodistribution des données afin de se protéger de sinistres.

Une des particularités d'OpenIO est que la firme a compris très tôt que le stockage objet pouvait être associé à des services de compute. La société a imaginé un framework serverless, baptisé Grid for Apps dont la vocation est de traiter des données lors de leur ingestion ou sur la base d'événements. Le premier marché visé est celui du stockage en masse d'images et de vidéos. La firme entend ainsi fournir un certain nombre de fonctions et d'API pour permettre l'indexation avancée des données (via des algorithmes de machine learning pour la reconnaissance de formes, de visages, le sous-titrage des films, etc.). Mais la technologie pourrait être étendue à de multiples domaines.

Aujourd'hui, le cœur du logiciel d'OpenIO est téléchargeable gratuitement, mais les connecteurs avancés ainsi



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

que l'interface d'administration graphique ne sont accessibles que pour les clients ayant souscrit au support de l'éditeur. Le support standard est proposé au prix de 5 centimes d'euro par an par Gigaoctet de données stocké. Il permet aux clients de choisir un pack de connecteurs parmi l'ensemble du catalogue offert par la firme. Le support Premium donne accès à tous les connecteurs, ainsi qu'à l'interface graphique d'administration. Il est offert au prix fixe de 90 000 \$ par an pour des capacités allant de 600 To à 5 Po. Le mode de tarification de Grid for Apps, encore en phase de développement avec plusieurs clients n'est pas encore défini.

■ SCALITY

Pionnier du marché mondial du stockage objet, Scality en est aujourd'hui l'un des leaders mondiaux. La firme a levé près de 152 M\$ de fonds en plusieurs tours de financement et compte parmi ses investisseurs HPE avec lequel la firme dispose d'une alliance stratégique (HPE revend la solution de Scality à son catalogue).

Fondé par Jérôme Lecat (CEO), Giorgio Regni (CTO), Brad King (Chief Engineering Architect), Daniel Binsfeld (VP Devops et support) et Serge Dugas (parti en 2017 chez Vade Secure), Scality a fait ses débuts en 2010. [Scality Ring](#), la solution de stockage objet phare de l'éditeur, est née du « spin off » de Bizanga Store, un système de stockage à grande échelle conçu à l'origine pour le

stockage d'e-mails, après le rachat de Bizanga par CloudMark en 2009.

De ces origines, Scality conserve de multiples intégrations avec les grandes plates-formes de messagerie électronique open source du marché comme DoveCot ou Zimbra, dont l'éditeur Synacor est un utilisateur historique de Scality Ring.

À la fin 2017, Scality comptait plus de 200 clients pour son logiciel de stockage objet phare, Scality Ring. Plusieurs de ces clients ont souscrit des licences au-delà de 100 Po. Outre sa flexibilité, sa résilience et ses performances, Scality Ring est réputé pour sa conformité à l'API S3, une compatibilité à laquelle la firme a porté énormément d'attention (le code du serveur S3 de Scality est par ailleurs un projet open source indépendant). L'essentiel de la R&D de la firme est basé à Paris, mais la firme dispose de bureaux à San Francisco, Tokyo, Londres, ainsi qu'en Allemagne et en Australie

AU-DELÀ DU STOCKAGE OBJET

Si Scality 2017 a basé sa croissance sur le succès de Ring, l'éditeur entend se diversifier dans la gestion de données multicloud avec son framework Zenko.io.

Zenko est une évolution [du serveur S3 open source de Scality](#) qui présente une API S3 unifiée en frontal de services de stockage courants comme Scality Ring, Microsoft

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Azure Blob, mais aussi de NAS NFS ou SMB. La solution inclut un contrôleur permettant de gérer des politiques sophistiquées de placement (et de déplacement) de données d'un service de stockage à un autre. Elle embarque aussi des fonctions de recherche avancées dans les métadonnées.

L'objectif de Zenko est de permettre aux entreprises de déployer des applications s'appuyant sur le protocole S3 tout en ayant la liberté de positionner leurs données à l'endroit de leur choix. Pour cela, le contrôleur se charge de la mise en œuvre de leurs politiques de placement de données. L'ensemble des données gérées par un contrôleur Zenko sont vues comme appartenant à un unique namespace. Le code de Zenko est open source, mais Scalality propose à ses clients une version supportée pour les entreprises.

■ ROZO SYSTEMS

Rozo Systems est une start-up française, née à l'origine dans la région nantaise sous le nom de Fizians. Elle propose une solution de [stockage NAS](#) distribuée à haute performance 100 % logicielle, qui a déjà séduit plusieurs grands utilisateurs dans le monde du calcul à hautes performances et des médias. RozoFS est ainsi utilisé par le CEO, l'institut Mario Negri à Milan ou par UMedia, l'une des grandes sociétés de production cinématographique européenne. La technologie est aussi mise en œuvre par plusieurs acteurs des effets spéciaux comme FuseFX ou

Muse VFX, deux sociétés qui ont travaillé sur les séries les plus récentes produites par Hollywood.

La technologie de Rozo Systems, actuellement en version 3.x, permet de constituer un système de stockage NAS massivement distribué (à la Isilon) pouvant agréger jusqu'à un maximum de 1024 nœuds de stockage. Ces nœuds peuvent être « hyperconvergés » et délivrer à la fois des services de stockage et de calcul, ou être dédiés au stockage, auquel cas ils délivrent des services NAS NFS et SMB à hautes performances, mais à un coût raisonnable.

UN ALGORITHME D'ERASURE CODING UNIQUE : LA MOJETTE

La technologie différenciante de RozoFS est un algorithme de code à effacement ([erasure coding](#)) basé sur la transformée Mojette (du nom d'un haricot blanc lingot populaire en Vendée et dans l'Ouest de la France).

Conçue à l'origine au sein du laboratoire CNRS [IRCCyN](#) de l'École Centrale de Nantes, cette transformation mathématique est une application discrète exacte de la [transformée de Radon](#). Appliquée aux systèmes de stockage, elle offre plusieurs atouts majeurs selon Rozo Systems. Le premier est celui de garantir une très haute intégrité des données sans impact majeur sur la capacité utilisée. Ainsi, d'après Pierre Evenou, on peut stocker cinq copies de la même information sur un cluster en ne



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

consommant qu'une fois et demie la capacité de l'information originale.

L'autre bénéfice de la Mojette est sa performance. Les algorithmes de code à effacement du marché s'appuient sur des variantes de codes Reed Solomon qui nécessitent des opérations d'inversion de matrices gourmandes en cycles CPU. La Mojette ne nécessite pas d'opérations complexes de multiplication et de division, mais de simples additions et soustractions et est donc deux à trois fois moins gourmande en CPU que l'un des codes d'écriture coding les plus optimisés du marché, l'[ISA-L d'Intel](#). Selon Rozo Systems sa technologie d'écriture coding affiche des performances environ 3 fois meilleures que celles d'Intel et jusqu'à 10 fois meilleures que celle d'un système de stockage objet comme Scality.

Selon la firme, la version 3 de RozoFS, lancée à l'automne 2017, apporte des performances en hausse de 60 % par rapport aux versions antérieures grâce notamment à des améliorations dans les communications réseau entre nœuds (support de RDMA over Ethernet) et à une optimisation du protocole SMB. Rozo Systems a aussi accéléré la performance des recherches de fichiers dans de grands systèmes de fichiers grâce à la mise à disposition d'une API spécifique.

Séparément, Rozo a noué des partenariats avec Dell

EMC, HPE, Lenovo et QCT pour supporter sa technologie sur les serveurs de ces constructeurs.

RozoFS est disponible dans une version communautaire gratuite téléchargeable sur GitHub sous licence GPL v2. L'algorithme d'écriture coding de cette mouture n'est toutefois pas optimisé, ce qui impacte les performances. La version commerciale, proposée par Rozo Systems et ses partenaires intègre quant à elle une version optimisée de l'algorithme et est pleinement supportée par l'éditeur.

■ PYDIO

Pydio est l'un des pionniers du partage et de la synchronisation de fichiers pour les entreprises. Il s'est peu à peu imposé comme une alternative « on premises » à des solutions SaaS comme DropBox, Box, OneDrive for Business, etc.

Sa solution a fait ses premiers pas en 2007 sous le nom AjaXplorer — un outil développé par Charles du Jeu, cofondateur et actuel CTO de la société — avant d'être rebaptisée Pydio en 2013.

À l'origine développé en PHP, le logiciel s'installe en frontal de multiples solutions de stockage (stockage local, partage réseau, serveur FTP ou stockage S3). Il est proposé dans une version communautaire gratuite sous licence libre AGPL et dans une version entreprise payante distribuée avec le plein support de la société.



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

PYDIO CELLS : UNE VERSION ENTIÈREMENT RÉÉCRITE DE PYDIO

Dans sa version la plus récente, baptisée Pydio Cells, le logiciel a été entièrement réécrit en langage GoLang et son interface a été repensée pour faciliter la collaboration entre groupes de travail. L'installation du logiciel a aussi été grandement simplifiée sur l'ensemble des OS supportés par l'éditeur (Linux, Mac OS, et VM sous linux).

Aujourd'hui rentable, Pydio compte près de 200 clients dans le monde pour sa version entreprise, explique David Gillard, le cofondateur de la société, soit un peu plus d'un million d'utilisateurs payants.

La plupart des clients, PME ou grands comptes, ont choisi la solution pour proposer des services avancés de partage de fichier et de collaboration sans avoir à dépendre de services [SaaS](#). Avec Pydio, ils peuvent en effet déployer ces services en mode "on-premises" (auto-hébergement), ce qui leur permet d'en conserver le plein contrôle.

Selon David Gillard, Pydio 8 et Pydio Cells vont cohabiter au sein du portefeuille de l'éditeur jusqu'en 2020. Mais l'avenir est clairement à Pydio Cells. D'ailleurs, un outil de migration automatisé de Pydio 8 vers Cells arrivera en septembre.

Pydio Cells reprend environ 90 % des fonctionnalités de Pydio 8 et en ajoute de nouvelles. Les principales améliorations fonctionnelles portent sur la gestion du "versioning" et sur les fonctions de collaboration — avec en particulier l'arrivée de fonctions avancées de chat dans les groupes de discussions.

D'autres aspects techniques du logiciel ont aussi été renforcés. La performance dans les grands environnements a été singulièrement améliorée - avec notamment un bien meilleur support des grands annuaires - de même que la gestion des fichiers volumineux. L'ensemble des fichiers gérés par Pydio est aussi désormais accessible directement via l'interface S3.

Selon David Gillard, le logiciel requiert également moins de mémoire vive que Pydio 8. Comme auparavant, le logiciel est accessible via une interface web sophistiquée ou via les clients mobiles pour IOS et Android.

■ OODRIVE

Connu initialement pour ses services de partage de fichiers et de collaboration en ligne (PostFiles) ainsi que pour ses services de sauvegarde en ligne, Oodrive a multiplié les acquisitions au cours des dernières années avec les rachats d'Active Circle, CertSign ou Orfea. Ces opérations lui ont permis de diversifier son offre dans la confiance numérique, le stockage de masse et l'archivage



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

ou la gestion des médias. Rentable, la firme [a levé 65 M€](#) en mars 2017 pour renforcer son portefeuille et accélérer son développement. Elle compte actuellement près de 400 salariés et réaliserait plus de 40 M€ de chiffre d'affaires auprès de 15 000 clients.

En quelques années, l'offre originelle de partage et de synchronisation de fichiers de l'éditeur s'est considérablement enrichie avec des offres adaptées à des besoins verticaux. Autour de l'historique PostFile, OoDrive a développé de multiples services comme iExtranet, une plate-forme collaborative sécurisée pour la gestion de contenus ou BoardNox, une solution sécurisée pour dématérialiser les conseils d'administration et autres réunions de gouvernance. A cela il faut rajouter aussi DilRoom, une solution de dataroom virtuelle pour permettre le partage sécurisé de documents durant une opération de fusion/acquisition ou durant la négociation de contrats sensibles.

Côté sauvegarde en mode SaaS, l'offre de backup des données des postes clients a, elle aussi, été modernisée et s'est progressivement étendue à la protection des serveurs ; même si le backup de serveurs avec la technologie de la firme n'est adapté de façon réaliste qu'à la protection des PME ou des agences.

ACTIVE CIRCLE : UN PIONNIER DU STOCKAGE DE DONNÉES EN VOLUME

Enfin, en matière de stockage, Oodrive a poursuivi le développement de la solution de stockage NAS en cluster d'Active Circle, [rachetée en 2014](#). La solution d'Active Circle est l'une des premières solutions de stockage NAS distribué apparue sur le marché mondial. OoDrive la positionne aujourd'hui sur le marché du stockage de données tièdes et froides et cible les gros producteurs de données qui ont besoin de conserver des centaines de téraoctets de données et plus.

Outre le monde du calcul scientifique, Oodrive cible le marché du stockage en masse de médias (images, vidéos). La solution compte actuellement quelque 150 entreprises clientes, dans le secteur de la vidéo (production TV, cinéma, postproduction) mais aussi dans le domaine de la gestion de grands fonds patrimoniaux (Assemblée Nationale, Museum d'Histoire naturelle, etc.) ou du stockage en masse d'image (Spatial, Géomapping...).

Plus récemment la société a aussi signé avec plusieurs grands comptes qui utilisent la solution pour le stockage des données de leurs départements communication. L'un des points forts de la solution est qu'elle est capable de gérer de multiples niveaux de stockage (disque rapide, disque lent, bibliothèques de bandes...), ce qui permet



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

d'optimiser le coût d'archivage des données tout au long de leur cycle de vie.

Il est à noter que la technologie d'Active Circle est aussi utilisée dans le cadre de la plate-forme technologique SaaS de l'éditeur. « Nous avons commencé à utiliser les technologies d'Active Circle pour nos besoins propres », explique ainsi Philippe Boyon en charge du marketing produit d'OoDrive. « Notre démarche est de faire converger les solutions du portefeuille vers une plate-forme commune sur laquelle reposeront toutes nos solutions ».

■ ATEMPO-WOOXO

L'année 2017 aura été chargée pour Wooxo, une start-up spécialisée dans la sauvegarde et la protection des infrastructures informatiques en mode on-premises et/ou en mode SaaS. La jeune société fondée par Luc d'Urso a en effet mis la main coup sur coup sur l'un des champions français historiques de la sauvegarde et de l'archivage, Atempo, avant de faire main basse sur Synerway, l'un des spécialistes français des appliances de sauvegarde.

L'objectif de la firme, fondée en 2010, est de créer un champion français et européen de la sauvegarde et de la protection de données à même de servir aussi bien les besoins des grands comptes que ceux des PME et des TPE. Wooxo s'est fait connaître avec sa solution Yoo Backup. Cette dernière est conçue pour permettre la sauvegarde

automatisée des PC et des serveurs d'une entreprise. Selon les options de déploiement retenues par les clients, Yoo Backup se déploie de multiples façons.

Le mode le plus simple est un déploiement sous forme 100 % logicielle avec une sauvegarde des données dans le cloud. Les données sont alors transférées dans deux datacenters redondants de Wooxo localisés en France. Une autre possibilité est le déploiement de la solution sous forme hybride. Le client installe alors une ou plusieurs appliances Wooxo dans ses datacenters et les données sont d'abord sauvegardées dans les appliances avant d'être répliquées vers le cloud de l'éditeur.

Enfin, l'entreprise peut déployer des appliances redondantes dans ces datacenters pour la sauvegarde de ses données et ne pas utiliser le cloud, s'assurant ainsi un plein contrôle de ses données. L'offre actuelle d'appliances de la firme permet de stocker de 1 To à 1 Po de données de sauvegarde.

ATEMPO VA PERMETTRE À WOOXO DE SE DÉVELOPPER SUR LE MARCHÉ DES GRANDS COMPTES

Si l'offre Yoo Backup cible principalement les PME, les solutions rachetées à Atempo visent le marché des grandes entreprises.



LES STARTUPS FRANÇAISES DU STOCKAGE

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Atempo n'est pas un nouveau venu sur le secteur. La firme est née il y a plus de 25 ans en France avant d'être absorbée par ASG Software en 2011. Après une passe difficile, ASG a finalement décidé de se reconcentrer sur son cœur de métier mainframe et a décidé de céder Atempo en 2017. C'est alors que Wooxo a mis la main sur la société.

Atempo apporte à Wooxo son logiciel de sauvegarde d'entreprise Time Navigator ainsi que sa solution d'archivage à long terme Digital Archive. Les deux offres comptent plus de 4500 clients dans le monde, dont plusieurs ayant plusieurs pétaoctets de données protégées par les deux logiciels. La solution Digital Archive a été retenue par de grands clients du secteur des médias et du calcul scientifique dans le monde. Atempo entend s'appuyer sur cette base installée pour séduire de nouveaux utilisateurs

confrontés à la problématique d'archivage de quantités massives de données.

Le dernier outil dans l'arsenal de Wooxo est l'offre d'appliances de backup de Synerway. La société veut s'appuyer sur le savoir-faire de Synerway pour développer une nouvelle génération d'appliances permettant de faire converger les technologies de Wooxo et Synerway pour servir le marché du mid-market, mais aussi celui des comptes publics européens.

Avec un CA combiné de plus de 20 M€, le groupe entend accélérer sa croissance et s'imposer comme l'un des champions européens de la protection de données. ■

CHRISTOPHE BARDY

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

LES OFFRES DE STOCKAGE EN CLOUD DES ACTEURS FRANÇAIS

Il n'y a pas qu'Amazon AWS, Microsoft Azure ou Google Cloud dans la vie. Plusieurs acteurs hexagonaux du cloud ont développé des services de stockage, qui sont autant d'alternatives possibles aux services proposés par les géants mondiaux. Et ils ont l'avantage de ne pas être soumis aux réglementations américaines intrusives telles que le Patriot Act ou le Cloud Act.

Par Christophe Bardy

FACE AUX GÉANTS AMÉRICAINS DU CLOUD, une poignée d'acteurs français de taille intermédiaire proposent aux entreprises une palette complète de services de stockage en cloud qui répliquent les caractéristiques essentielles des services américains ou offrent des caractéristiques différenciantes.

Parmi ces acteurs, nous avons passé en revue les offres d'OVH, Online, Orange Business Services et Outscale, quatre acteurs qui ont la particularité d'avoir développé des offres de cloud à l'échelle internationale (quoiqu'à des échelles différentes).

De ces quatre fournisseurs, OVH est sans doute celui qui a eu la politique de développement la plus agressive au cours des trois dernières années, même si la croissance de la société n'a pas forcément toujours été de tout repos. Orange Business Services a lui aussi dû affronter son lot de



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

turbulence. Après avoir misé sur le cloud souverain avec CloudWatt — et largement échoué —, en complément de son offre historique de cloud privé VMware, la division entreprise d'Orange a finalement lié son avenir dans le cloud à celui de Huawei, dont elle a adopté la technologie cloud pour bâtir son cloud public à l'international.

Grand acteur français historique de l'hébergement de serveurs, Online a commencé avec des objectifs modestes et n'a que récemment accéléré son développement dans le cloud via sa marque Scaleway. La particularité de l'offre du fournisseur est qu'elle s'appuie sur des briques matérielles et logicielles largement développées en interne. Une stratégie qui a produit des services originaux comme l'offre de stockage de données à long terme et à très bas coûts C14.

Enfin, Outscale, désormais intégré au groupe Dassault, poursuit son développement avec une stratégie originale, qui s'appuie sur des briques technologiques du marché (serveurs et équipements réseau Cisco, stockage NetApp). Celle-ci est orchestrée par une technologie cloud développée en interne et propose des services dont les API sont compatibles avec ceux d'Amazon AWS (quoiqu'avec un nombre de services bien inférieur).

Dans cet article, nous nous sommes intéressés en particulier au volet stockage des offres de ces 4 fournisseurs, dont nous vous proposons un tour d'horizon.



OVH.com

■ OVH

Outre ses services de stockage primaires ([lire le numéro 14 de Storage](#)), OVH propose actuellement deux services de stockage cloud à ses clients. Le premier baptisé OVH Object Storage est une solution de stockage objet qui s'adosse à OpenStack Swift. Déployée dans chaque région cloud d'OVH, le service est conçu comme une alternative à AWS S3. Il offre toutefois plusieurs caractéristiques tarifaires intéressantes, à commencer par un prix unique au Go stocké (0,01 € par gigaoctet et par mois) et une absence de tarification additionnelle à la transaction. Le seul coût additionnel à prévoir est facturé lors de l'extraction de données hors du cloud OVH (0,01 € par gigaoctet). Selon la société, chaque donnée écrite dans le cloud OVH est répliquée trois fois afin de se protéger contre d'éventuelles défaillances matérielles. OVH garantit un taux de disponibilité de 99,9 % et une durabilité des données de 100 %.

En parallèle de ce premier service de stockage, le fournisseur propose un service d'archivage de données baptisé OVH Cloud Archive. Celui-ci est considéré comme une alternative à AWS Glacier. Le service s'appuie là encore

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

sur OpenStack Swift, mais il est aussi accessible via une multitude d'autres protocoles comme FTP, SFTP et rSync. Depuis un client Linux, il est également possible de monter un « volume » Swift comme un système de fichiers via [le système de fichiers virtuel SVFS](#). Le stockage de données dans OVH Cloud Archive est facturé 0,002 € par Go et par mois. Une surcharge est aussi prélevée à l'ingestion et à l'extraction de données (0,0115 € par Go).

À titre de comparaison, le stockage de données dans AWS Glacier est facturé 0,004 \$ par Go et par mois, tandis que l'extraction est facturée entre 0,01 \$ (mode standard) et 0,03 \$ (mode accéléré) par Go. À cela s'ajoutent des frais par transaction et des coûts de transferts de données sortantes sur le réseau. Autant dire que l'offre d'OVH est largement compétitive d'un point de vue tarifaire avec celle d'AWS. Notons pour terminer que le temps de désarchivage varie entre 10 mn et 12 heures.



■ ONLINE

Le fer de lance de l'offre d'Online en matière de stockage cloud est le service d'archivage de données à long terme C14 (pour carbone 14). Présenté comme un service

de coffre-fort numérique, C14 a été spécifiquement conçu pour délivrer des services de stockage à long terme à un coût plancher. Le service s'appuie sur une pile de stockage logicielle développée par Online qui met en œuvre un mécanisme de code à effacement maison pour la distribution des données au sein du cluster.

Le déversement des données dans C14 s'effectue pour l'instant via des protocoles comme FTP, SFTP, Rsync ou SCP, mais il devrait prochainement s'élargir à S3. La durabilité des données est de « 11 neufs » pour les classes de services standard et intensives. Elle approche les « 12 neufs » pour la classe de service Entreprise. Le désarchivage de données requiert un temps d'attente de 2 heures, après lequel Online garantit un débit de 650 Mo/s en transfert (si la liaison du client le permet). Les données du service sont répliquées entre les datacenters DC2 (Ivry) et DC4 (Paris 15e), ce dernier étant situé dans un ancien bunker souterrain antiatomique de la défense passive. Celui-ci a été réaménagé en datacenter en 2016 et dispose d'une capacité d'hébergement de 9000 m2.

Côté tarifaire, le prix le plus bas pour le stockage de données est de 0,002 € par Go et par mois, avec un coût de 0,01 € par opération de désarchivage (offre standard). L'offre Entreprise facture le stockage à 0,004 € par Go et par mois et les opérations à 0,025 € tandis que l'offre intensive ne facture que le stockage (0,005 € par Go et par

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

mois). Avec de tels tarifs, C14 est entre 20 % et six fois moins coûteux que les services d'Amazon, selon l'offre retenue. Le service est également bien moins coûteux que Google. Selon l'offre et la durée de rétention retenue, il est également moins cher que l'offre d'archivage d'OVH qui, du point de vue des prix, est son concurrent le plus proche.



Business Services

■ ORANGE BUSINESS SERVICES

L'opérateur de cloud français propose dans le cadre de son Cloud Flexible Engine, une offre baptisée « stockage objet » s'appuyant sur la technologie développée par Huawei. Compatible avec le protocole S3, le service est disponible dans toutes les régions où l'opérateur a déployé son cloud (Paris pour l'Europe, Singapour pour l'Asie et Atlanta pour l'Amérique du Nord).

Orange Business Services positionne son offre de stockage objet pour le « stockage et la distribution sur Internet de contenus variés (données, fichiers, multimédia), pour l'hébergement de sites statiques et pour des besoins de sauvegarde et d'archivage ».

L'opérateur garantit une disponibilité du service de 99,9 % et une durabilité de « 9 neufs ». Les tarifs pour le stockage de données vont de 0,0175 € par Go et par mois à 0,0191 € par mois selon la quantité de données stockées sur le service. À ce prix s'ajoutent des coûts de trafic sortant allant de 0,038 € à 0,070 € par Go transféré, ainsi qu'une facturation à la transaction (GET/PUT/LIST) allant de 0,0036 € par tranche de dix mille. Le détail complet des tarifs est consultable sur [la fiche tarifaire publiée par l'opérateur](#).



■ OUTSCALE

En plus de son offre de stockage en mode bloc, Outscale, la division cloud de Dassault, propose une offre de stockage objet compatible S3, baptisée Object Storage Unit (OSU). S'appuyant sur la technologie NetApp et certifié Altavault, le service peut être utilisé en complément de baie NetApp « on premise » comme service de débordement pour le tiering dans le cloud de données froides.

La solution est disponible en deux versions, entreprise et premium. Elles se différencient par le facteur de



LES OFFRES DE STOCKAGE EN CLOUD DES ACTEURS FRANÇAIS

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

réplication utilisé pour les données (trois fois sur un site ou six fois sur deux datacenters différents). Dans les deux cas, le service garantit une durabilité des données de « 12 neufs », soit le niveau le plus élevé de tous les services de stockage objet français.

En Europe, le service premium facture le gigaoctet de données stocké à 0,04 € par mois, un prix auquel s'ajoute une facturation de 0,08 € par gigaoctet de données sortant (l'ingestion de données n'est pas facturée). Une offre avec trafic sortant illimité est proposée au prix forfaitaire de 0,1 € par Go stocké et par mois.

Le service Entreprise facture quant à lui le gigaoctet de données stocké à 0,1 € par mois. A cela s'ajoute la même

facturation de 0,08 € par gigaoctet de données sortant que pour le service premium. Une offre avec trafic sortant illimité est proposée au prix forfaitaire de 0,16 € par Go stocké et par mois.

Il est à noter qu'Outscale ne facture aucun frais par opération (Get, Put, List) contrairement à la plupart des services de stockage objet des grands clouds américains. La liste complète des tarifs de l'opérateur cloud est [consultable en ligne](#). ■

CHRISTOPHE BARDY

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

LES 8 START-UPS US DU STOCKAGE À SUIVRE

De nouveaux acteurs innovants tentent de se faire une place face aux acteurs traditionnels dans le monde du stockage objet, du NAS Scale-out et du stockage SAN.

Par Christophe Bardy

LES DERNIÈRES ANNÉES ont vu apparaître une nouvelle génération de start-ups de stockage qui tentent de bousculer l'ordre établi en misant agressivement sur des technologies émergentes. Dans le monde du stockage objet, Minio a fait le pari du devops en mariant étroitement sa technologie de stockage objet compatible S3 avec un [framework](#) Lambda et en simplifiant son déploiement dans les environnements conteneurisés. Dans le monde du stockage NAS, les nouveaux venus comme WekaIO ou Elastifile, misent quant à eux sur la performance, la simplicité de déploiement et les interactions avec le cloud.

Ils entendent séduire une nouvelle catégorie de clients, dont les architectures actuelles ne suffisent plus pour faire face à l'accroissement des besoins des applications. Dans le monde du stockage SAN, des acteurs comme E8 Storage, Excelero ou Vexata ont fait le pari de la performance pure en développant des architectures capables de délivrer des millions d'IOPS avec une latence très faible.

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Ces architectures sont conçues pour répondre aux besoins des nouvelles applications webscale, des applications d'IA et de machine learning, ainsi qu'à ceux des bases de données à très hautes performances. Enfin, le petit dernier de la liste, DriveScale, a fait le pari de la désagrégation et des architectures composables en proposant une technologie matérielle et logicielle permettant d'allouer dynamiquement des disques durs à des serveurs dans un rack à partir de [JBOD](#) mutualisés.



Basé en Californie à Palo Alto, Minio a été co-fondée par Anand Babu Periasamy, l'un des principaux développeurs et ex-CTO de [Gluster](#). La société s'est fait un nom en développant une solution de stockage objet Open Source [compatible S3](#), qui vise les développeurs d'applications. Alors que les autres acteurs du stockage objet se concentrent sur le marché du stockage de masse, Minio a mis l'accent sur les performances, la résilience et sur l'environnement de développement autour de sa solution.

.....

CRÉÉE EN NOVEMBRE 2014, Minio compte moins de 20 salariés et a levé un total de 23,3 M\$ auprès des fonds AME Cloud Ventures, Dell Technologies Capital, General

Catalyst Partners, Index Ventures et Nexus Venture et auprès d'investisseurs privés dont Mark Leslie (CEO historique de Veritas), Jeff Rottschild (cofondateur de Veritas), Ben Golub (CEO de [Docker](#)) ou Marten Mickos (ex-MySQL).

Minio peut se déployer comme un exécutable traditionnel sous Windows, Linux, MacOS ou FreeBSD. Mais dans 90 % des cas, il est mis en œuvre sous la forme d'un conteneur Docker. Idéalement, explique Anand Babu Periasamy, l'objectif est que les développeurs déploient programmatiquement leurs applications en même temps que Minio.

Le logiciel dispose de deux modes de fonctionnement. Le premier permet de mettre en place un stockage compatible S3 au-dessus de volumes classiques en mode bloc ou d'un service de stockage en cloud. Minio présente alors les volumes de stockage consommés comme un stockage S3. Dans ce mode, Minio permet à un développeur de disposer d'un stockage objet à peu près n'importe où, y compris sur sa machine de travail.

Le second mode de déploiement, bien plus sophistiqué, est baptisé Distributed Minio. Dans ce mode, Minio est capable d'agréger la capacité de multiples volumes résidant sur un ou plusieurs serveurs physiques ou virtuels en un espace de stockage objet protégé par un mécanisme

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

de [code à effacement](#) sophistiqué.

Ce code tire notamment parti des dernières capacités des puces x86 et il est optimisé pour les nouvelles unités de calcul en virgule flottante des puces Intel et AMD ce qui lui confère des performances élevées. L'algorithme d'effacement de Minio est couplé à un mécanisme anti "bit-rot" qui permet au logiciel d'assurer l'intégrité des données en cas de défaillance d'éléments de l'infrastructure ou de corruption silencieuse des informations stockées sur les disques. Ces caractéristiques en font un stockage adapté à un grand nombre d'applications y compris lorsqu'elles sont relativement sensibles à la latence.

Le [stockage objet](#) ("object store") est compatible avec le protocole S3 et accepte l'écriture de tout type d'objets tant que leur taille unitaire ne dépasse pas 5 To. Point intéressant pour les développeurs, Minio supporte aussi les mécanismes de notifications Lambda d'AWS - ce qui permet à certains événements de déclencher des appels à des fonctions spécifiques. Cela peut par exemple aller de la conversion automatique d'une image ou d'une vidéo après ingestion, à la mise en œuvre de mécanisme de scoring ou de filtrage de données.

Minio est téléchargeable et utilisable gratuitement et l'éditeur propose une palette de prestations de support payantes destinée aux grands clients.



ELASTIFILE

Elastifile est une startup d'origine israélienne qui a conçu un système de fichiers distribué qui permet de combiner les atouts du stockage 100 % Flash local et du stockage en cloud.

Parmi les fondateurs de la firme, on compte notamment Shahar Frank - le cofondateur et "chief scientist" de XtreamIO et ex-architecte en chef d'Exanet (qui a servi de base au Fluid File System de Dell) - et Roni Luxenburg, l'ex-directeur de l'ingénierie de Qumranet (créateur de l'hyperviseur KVM). L'équipe comprend également Avraham Poza Meir, l'ex-CTO d'Anobit (racheté par Apple) et VP ingénierie de SanDisk.

LA SOCIÉTÉ DISPOSE de bureaux à Santa Clara, dans la Silicon Valley, et à Herzliya, au nord de Tel-Aviv. Elle compte parmi ses actionnaires des acteurs de poids comme Dell Technologies, Cisco, Lenovo et Western Digital. Elle emploie actuellement quelque 100 salariés et entend en embaucher une quinzaine de plus et comptait plus d'une cinquantaine de clients dans le monde au début de l'année.

[Le système de fichiers distribué ECFS d'Elastifile](#) s'installe sur des nœuds x86 standard et il est à même de créer des

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

pools de stockage flash résilients en agrégeant la capacité offerte par chacun d'entre eux.

Contrairement à de nombreux systèmes de fichiers distribués, ECFS est optimisé pour délivrer une faible latence locale et donc des performances maximales, le tout sans sacrifier les services. Il s'appuie pour cela sur un algorithme de consensus maison, baptisé Bizur, conçu pour optimiser la latence des opérations de stockage. Le système de fichiers d'Elastifile met aussi en œuvre des mécanismes de déduplication globale et de compression de données pour optimiser la capacité de stockage et minimiser l'usure des SSD et il intègre un mécanisme de gestion de la QoS (qualité de service) avancé.

Elastifile a aussi été conçu pour interagir de façon optimisée avec les services de stockage en cloud. ECFS implémente ainsi un système de tiering baptisé CloudConnect, qui permet de transférer les données peu utilisées hors du [file system](#) distribué pour les stocker sur un service ou un système de stockage objet (compatible S3). Cette fonction de tiering automatisé vers le cloud a été pensée pour abaisser le coût de stockage des données dormantes. De façon intéressante, Cloud Connect ne s'appuie pas sur un format de données propriétaire, ce qui fait que les données écrites sur le système objet peuvent être accédées de façon transparente via le protocole S3.

Le logiciel d'Elastifile est vendu par abonnement à partir de deux cents par gigaoctet et par mois, un coût auquel il faut ajouter celui des serveurs et des SSD ou celui des instances cloud et du stockage flash en cas de déploiement dans le cloud. Il est à noter qu'ECFS est proposé sous la forme d'un service NAS prêt à l'emploi dans la marketplace de Google, Google Cloud Launcher, et qu'il est aussi possible de déployer le logiciel sur Amazon AWS et Microsoft Azure.



WEKAIO

Né à Tel-Aviv, mais avec un siège basé à San Jose, WekaIO a conçu un système de fichiers NAS distribué baptisé Matrix, qui promet de concilier hautes performances, résilience et faible coût de stockage. La société, qui a notamment noué un partenariat avec la division HPC de HPE, vise le marché du calcul à hautes performances et du stockage pour les applications d'IA et de machine learning.

WekaIO a été fondé par d'anciens responsables de XIV comme Omri Palmon, le Chief Product Officer, Liran Zvibel, le CTO, ou Aaron Ben Dayan, le chief architect. La firme a levé environ 32,5 M\$ auprès des fonds Gemini Israel Ventures, Norwest venture partners, Qualcomm Ventures et Walden Riverwood Ventures.



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

MATRIX EST UN SYSTÈME DE FICHIERS distribué conçu pour tirer parti de la performance offerte par le stockage flash local embarqué sur les serveurs, tout en exploitant le faible coût des systèmes de stockage objet. Doté de capacités natives de tiering, le filesystem de Weka peut ainsi délivrer des performances extrêmement élevées pour les données chaudes, tout en fournissant un coût de stockage faible pour les données froides et en s'appuyant sur des services cloud comme Amazon S3, Google Cloud Storage ou sur des systèmes de stockage objet on premises de type Cloudian, HGST, Scality.

UN SYSTÈME DE FICHIERS MODERNE CONÇU POUR TIRER PARTI DE LA FLASH

Pour tenir ses promesses en termes de performances, Matrix s'appuie sur une pile réseau optimisée pour réduire la latence et pour transmettre plus efficacement les commandes de stockage aux périphériques Flash à travers le réseau. « En produisant ainsi notre propre système temps réel et en opérant notre propre pile réseau et notre propre pile de gestion d'entrées/sorties, nous avons permis au [software defined storage](#) de délivrer de vraies performances » explique Liran Zvibel.

WekaIO permet d'agréger la capacité des SSD installés dans des serveurs pour la présenter comme un pool partagé à très hautes performances aux applications installées sur des serveurs. Les données stockées sur le cluster

sont accessibles via les protocoles NFS, SMB ou HDFS. La résilience est assurée par la mise en œuvre d'un algorithme d'[erasure coding](#) maison conçu pour maximiser les performances et minimiser les temps de reconstruction de données en cas de défaillance. MatrixFS propose ainsi deux modes de protection de données, un mode N+2 tolérant la défaillance de deux périphériques pour chaque donnée stockée et un mode N+4.

Point intéressant, Matrix ne se limite pas aux seuls SSD. Il peut aussi agréger au sein d'un même namespace, la capacité fournie par les SSD et celle délivrée par des systèmes de stockage objet, on ou off-premises. La gestion du tiering entre les deux classes de stockage se fait à base de règles personnalisées ou sur une base automatique (le système appliquant alors ses règles par défaut).

PLUSIEURS OPTIONS DE DÉPLOIEMENT POSSIBLES

MatrixFS nécessite au minimum six serveurs pour démarrer et peut être utilisé sur des clusters allant jusqu'à environ 1000 nœuds (limite certifiée à ce jour). Mais selon la firme, il serait possible de déployer le logiciel sur un maximum de 20 000 nœuds contribuant au stockage. Un cluster WekaIO peut également supporter jusqu'à 100 000 nœuds, si les autres machines ne font que consommer du stockage sans contribuer de capacité. Les nœuds peuvent être physiques ou virtuels, on premise ou dans le cloud et il est aussi possible de déployer le logiciel sous forme de conteneurs Docker.

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Il est possible d'affiner de façon très pointue la performance d'un cluster. Par défaut, MatrixFS se voit allouer un cœur CPU sur chaque nœud contribuant au cluster pour les opérations de stockage. Mais il est possible d'allouer plus de cœurs aux processus de stockage, la performance IOPS évoluant alors linéairement avec le nombre de cœurs alloués.

UN ACCENT MIS SUR LES UTILISATEURS AYANT BESOIN DE BEAUCOUP DE CAPACITÉ ET DE PERFORMANCES

WekaIO vise, dans un premier temps, le marché des utilisateurs ayant besoin de beaucoup de performance à un coût raisonnable. WekaIO revendique des performances très supérieures à celles de Spectrum Scale (ex GPFS) d'IBM et une plus grande simplicité de mise en œuvre. Le logiciel est aussi une alternative à de grandes installations NAS à base de nœuds Isilon, NetApp ou Panasas.

WekaIO Matrix est vendu aux environs de 1000 \$ par To lorsqu'il est déployé sur des serveurs dédiés. En mode hyper-convergé, le logiciel est facturé 10 000 \$ par nœud avec jusqu'à deux cœurs dédiés aux processus Matrix et une capacité illimitée de SSD. Ce prix passe à 12 500 \$ par nœud fournissant du stockage avec un plafond de huit cœurs dédiés aux processus de stockage (ce tarif est lié au fait que la performance de Weka augmente quasi linéairement avec le nombre de cœurs alloués au logiciel).



VEXATA

Fondé par d'anciens dirigeants d'EMC et de VMware, Vexata a conçu une architecture de stockage distribuée qui permet de tirer parti au mieux des performances offertes par les disques NVMe sans contraindre les entreprises à revoir en profondeur leur infrastructure SAN.

CETTE ARCHITECTURE S'INCARNE pour l'instant dans [la baie SAN Fibre Channel VX-100](#), un système de stockage en châssis de 6U qui accueille jusqu'à 16 lames de stockage équipées chacune de 4 SSD NVMe (soit 64 disques par baie). Ces lames sont en fait des microserveurs motorisés par des processeurs ARM Cavium, dont la mission est d'extraire le maximum de performance des [SSD NVMe](#) tout en délivrant une latence optimale.

Les lames sont interconnectées entre elles via un backplane redondant Ethernet. Sur ce backplane viennent aussi s'insérer deux contrôleurs redondants à base de puces Xeon qui exécutent les services avancés de la baie (thin provisioning, snapshot, réplication, clones). Ces contrôleurs, motorisés par l'OS maison, VX-OS, jouent le rôle d'un plan de contrôle distribué. Ce plan de contrôle est accéléré par une série de FPGA présents sur les



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

commutateurs Ethernet redondants de la baie.

L'architecture de la firme permet selon Surya Varanasi, son CTO, d'assurer une latence minimale. La pile de gestion des I/O de la baie aurait ainsi une latence de l'ordre de 5 micro-secondes, contre 250 à 1250 microsecondes pour la plupart des architectures flash traditionnelles.

RÉPONDRE AUX BESOINS DE PERFORMANCES DES APPLICATIONS MODERNES

En misant sur une architecture logicielle et matérielle distribuée, Vexata promet des performances très élevées, même pour les applications les plus exigeantes. Comme l'explique Surya Varanasi, la plupart des baies actuelles s'appuient sur un large étage de cache pour délivrer des performances élevées. Si cette stratégie s'avère très efficace pour des applications qui effectuent essentiellement des opérations de lecture, elle s'effondre lorsque la proportion d'écritures et de lectures aléatoires s'accroît.

Le problème, selon Vexata, est que la plupart des workloads analytiques modernes, ainsi que les applications émergentes de machine learning ou de deep learning ont des profils I/O radicalement différents des applications historiques, avec un ratio lecture/écriture qui peut approcher les 50/50. Dans un tel scénario, la performance des architectures de stockage historiques s'effondre.

C'est pour faire face à cette évolution des applications que Vexata a conçu son architecture, explique le CTO. Selon lui, une baie Vexata maintient sa performance, quel que soit le type de profil d'I/O des workloads auxquelles elle est confrontée.

Avec une application OLTP Oracle (sous Oracle RAC avec 8 nœuds), la baie délivrerait 6 millions d'IOPS avec une latence de 0,4 ms contre 275 000 IOPS à 0,9 ms pour son meilleur concurrent traditionnel. Avec SAS Kx, Vexata annonce pouvoir soutenir une bande passante de 45 Go/s avec 240 sessions en parallèle (contre 4 Go/s et 20 sessions pour son meilleur concurrent).

Cette performance peut être encore plus élevée avec la version de la baie utilisant des disques Intel Optane, la VX-100M. Les analystes d'ESG ont ainsi fait fonctionner en parallèle un test HammerDB (simulant un workload OLTP) pendant qu'une requête OLAP consommait 21 Go/s de bande passante sur la baie. Le résultat est étonnant avec un score de 8,5 millions d'IOPS et un temps de réponse moyen rapporté par Oracle Enterprise Manager de 12 à 30 microsecondes.

Vexata propose aussi une configuration baptisée VX-100FS qui associe 4 contrôleurs Spectrum Scale (ex-GPFS) à sa baie VX-100 dans une configuration dédiée au stockage de données non structurées (support de GPFS,

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

NFS et SMB). Selon le constructeur, cette solution peut soutenir plus de 1 million d'opérations fichiers par seconde et délivrer un débit soutenu de 50 Go/s en lecture, 20 Go/s en écriture (30 Go/s en usage mixte). Des performances qui en font une alternative séduisante à des systèmes comme le FlashBlade de Pure Storage pour les applications intensives d'AI et de machine learning.

UN PRIX TRÈS RAISONNABLE ET UN SUPPORT NATIF DU FIBRE CHANNEL

Outre ces performances élevées, Vexata met en avant plusieurs points forts de son architecture. Le premier est que la baie est une baie en attachement SAN Fibre Channel. Elle s'insère donc de façon transparente dans des architectures de stockage existantes de grands comptes.

Second point fort, Vexata a serré les prix. Une configuration de base avec 4 lames de stockage est ainsi proposée à partir de 75 000 \$, tandis que la déclinaison la plus musclée avec 16 lames et 64 SSD (environ 200 To de capacité brute et 155 To de capacité utile) est proposée à un prix public inférieur à 500 000 \$. Le prix de la version Optane de la baie n'a pas été précisé.

UNE VERSION « SOFTWARE-DEFINED » EN PRÉPARATION POUR 2018

Vexata n'entend pas en rester là. Si la firme propose une appliance prête à l'emploi avec le VX-100, elle travaille

d'arrache-pied à une version 100 % logicielle de sa technologie. L'idée est de pouvoir offrir aux clients qui le souhaitent la possibilité de mettre en œuvre la technologie de la firme sur un cluster de serveurs de stockage NVMe banalisés. Seule petite réserve, les nœuds contrôleurs continueront sans doute au moins dans un premier temps, à être délivrés par Vexata.



E8 STORAGE

Fondé par d'anciens responsables de XIV, dont Zivan Ori, ex-CTO de XVI chez IBM, E8 Storage a conçu une architecture de stockage NVMe over fabrics capable de délivrer des performances hors normes.

E8 Storage a été fondé en 2014 à Tel-Aviv et a désormais son siège à Santa Clara dans la Silicon Valley. La firme a levé 18 M\$ en deux tours de financement auprès des fonds Vertex Ventures, Magma Ventures et Accel Partners.

LA TECHNOLOGIE DE LA FIRME est proposée sous la forme d'appiances prêtes à l'emploi, les baies de la série D24, et [sous forme 100 % logicielle](#). Elle s'appuie sur une architecture distribuée originale, qui associe des systèmes de stockage Flash bicontrôleurs de haute performance pour



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

la persistance des données, à une série d'agents logiciels installés sur les serveurs accédant à ce stockage.

Dans cette architecture originale, la gestion du plan de contrôle est partagée entre le système de stockage et les agents distribués sur les serveurs hôtes (Linux uniquement). Le nœud de stockage implémente une fonction de coordination des fonctions de contrôle des agents via un serveur de métadonnées (qui peut lui-même être distribué entre plusieurs baies de stockage). Chaque agent consomme quant à lui un ou plusieurs cœurs sur son serveur hôte (configurable par l'administrateur) et est responsable d'une partie des fonctions de gestion du plan de contrôle de la solution d'E8.

De facto, les agents agissent comme un plan de contrôle distribué (96 serveurs hôtes max) pour la plate-forme de stockage centralisée à hautes performances. Chaque agent a aussi en charge la gestion des chemins de données, une fonction traditionnellement gourmande en ressources et qui du fait de sa distribution sur les serveurs hôtes permet de maximiser la performance en déchargeant les contrôleurs de la baie.

Il est à noter que les agents des serveurs hôtes sont « stateless » et ne communiquent pas entre eux, ce qui fait qu'une défaillance d'un serveur hôte n'affecte pas le fonctionnement de l'architecture.

BASCULE VERS UN MODÈLE 100 % LOGICIEL

Initialement, E8 Storage avait fait le choix de packager sa technologie avec sa propre appliance de stockage, un nœud bicontrôleur équipé de 24 emplacements pour disques NVMe à deux ports fabriqué par l'OEM taiwanais AIC. Mais depuis peu, E8 propose sa technologie sous forme 100 % logicielle.

Selon Julie Herd, en charge du marketing produit de la société, E8 Storage a vu se multiplier les offres de serveurs capables de jouer le rôle tenu actuellement par le nœud AIC : « Des constructeurs comme Celestica ou WD ont désormais leur propre JBOF ("Just a bunch of Flash") bi-contrôleur ». E8 Storage a aussi validé plusieurs serveurs HPE, Dell EMC et Lenovo à base de puces Xeon Scalable Platform (ex « Skylake-EP ») pour jouer le rôle de nœuds de stockage. Parmi ceux-ci figurent [le Dell EMC PowerEdge R740](#), le Lenovo ThinkServer SR630 et [les serveurs HPE Proliant 360 et Proliant 380 Gen10](#).

Cette multiplication des configurations va accroître le choix des entreprises selon la firme. Ces dernières vont pouvoir faire le choix entre plusieurs types d'architectures. Pour celles qui recherchent la résilience maximale, il sera possible de continuer à utiliser des JBOF bi-contrôleurs équipés de disque NVMe à deux ports. Cette configuration protège d'une panne de contrôleur au niveau du nœud de stockage et vient compléter la protection Raid-6

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

logicielle fournie par la couche de contrôle distribuée dans les serveurs hôtes.

Pour la majorité des clients, il sera possible d'utiliser l'aptitude des agents logiciels d'E8 Storage à piloter plusieurs nœuds de stockage pour obtenir une résilience similaire, mais en s'appuyant sur plusieurs nœuds de stockage mono contrôleur plus économiques, équipés de disques NVMe à un seul port (moins chers et disponibles auprès d'un plus grand nombre de fournisseurs). Dans ce cas, chaque agent écrit chaque donnée vers deux cibles de stockage, ce qui fournit la protection contre une défaillance de nœud de stockage. Dans ce mode, E8 Storage met aussi en œuvre un mécanisme Raid-5 logiciel pour se protéger des défaillances individuelles de disques.



EXCELERO

Fondé en 2014 à Tel-Aviv et basée comme E8 Storage à Santa Clara, Excelero a développé une architecture de stockage distribué à très hautes performances conçue pour tirer parti des capacités des disques SSD NVMe et de la performance des réseaux Ethernet modernes. Baptisée NVMesh Server SAN, la technologie de la société s'appuie sur une pile logicielle maison permettant de délivrer des millions d'IOPS avec une latence extrêmement élevée. La firme vise dans un premier

temps le marché du HPC, des bases de données à hautes performances, celui des applications webscale ainsi que le marché des médias et du cinéma.

NVMESH SERVER SAN est une technologie logicielle, qui permet de créer des pools de stockage en unifiant les multiples disques SSD NVMe présents sur les serveurs d'un cluster. Dans la pratique, Excelero installe un pilote en mode bloc propriétaire sur chacune des machines Linux d'un cluster. Ce pilote permet à chaque machine d'être serveur ou client de stockage ou les deux en même temps. « Nous vous permettons de partager des disques NVMe à l'échelle de votre datacenter », explique le CTO de la firme Yaniv Romem. « Une application tournant sur un serveur A peut consommer du stockage sur les serveurs B, C et D ».

UN ACCÈS À DES DISQUES NVME DISTANTS VIA UN PROTOCOLE PROPRIÉTAIRE SUR RDMA

Des datacenters utilisent la technologie d'Excelero à grande échelle pour assembler des environnements de type SAN à partir de serveurs standards, de cartes réseau RDMA et de disques NVMe. Parmi les premiers clients de la firme figurent la division Predix de General Electric, Hulu, la NASA et PayPal. D'autres utilisent la technologie pour délivrer un étage de cache à très hautes performances. C'est le cas de SciNet qui utilise NVMesh Server SAN dans le rôle de Burst Buffer pour son stockage



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Spectrum Scale (ex-GPFS). L'organisme HPC canadien a ainsi assemblé un cluster NVMe de 10 nœuds avec 80 SSD NVMe qui délivre une bande passante agrégée de 148 Go/s en écriture et de 230 Go/s en lecture et peut supporter jusqu'à 20 millions d'IOPS (lecture aléatoire de blocs de 4K).

NVMe Server SAN s'appuie sur une technologie maison baptisée RDDA ([Remote Direct Drive Access](#)) pour permettre à tout serveur de lire et écrire sur n'importe quel disque NVMe. RDDA s'appuie sur une « fabric » Ethernet RDMA (Remote Direct Memory Access) et permet de transmettre des commandes RDMA au-dessus d'un réseau sans impact sur la performance processeur. Selon la firme, cela lui permet de supporter la montée en charge de façon linéaire.

RDDA s'appuie sur RoCE ([RDMA over converged Ethernet](#)) pour relier tous les disques d'un cluster. « La technologie fait croire aux cartes réseau qu'elles écrivent en mémoire alors qu'elles écrivent en fait sur un disque NVMe », explique Romem. Selon Excelero, RDDA permet des communications directes en mode point à point entre le serveur « client » et les disques cibles sur les autres nœuds. Ce faisant, RDDA contourne la pile NVMe over fabrics et permet l'économie de quelques dizaines de nanosecondes de latence. D'après la firme, la différence entre un accès disque local NVMe et un accès distant à

un disque NVMe via RDDA n'implique qu'une pénalité de quelques nanosecondes, en fait la latence du réseau entre les deux nœuds.

NVMe Server SAN met aussi en œuvre une séparation claire entre plan de contrôle et plan de données. Le plan de contrôle, ou Topology Manager (TOMA) est implémenté de façon distribuée, en parallèle aux pilotes installés sur chaque serveur. Il s'exécute sous la forme d'un module tournant en mode « user space » afin de minimiser la latence des communications. Ce plan de contrôle permet de détecter les anomalies (perte de lien ou de disque) et de contourner les éventuelles défaillances.

NVMe peut fonctionner en mode convergé, désagrégé ou en mode hybride. Le produit supporte aussi les SSD SAS et SATA, mais sans les bénéfices de performance liés à l'usage de NVMe. Côté protection de données, NVMe supporte les modes Raid 0 (agrégation par bande), Raid 1 (miroir) et Raid 10 (agrégation de grappes de disques répliquées). Les données sont distribuées entre de multiples SSD sur de multiples serveurs. L'architecture détecte des pannes à l'échelle des serveurs et du rack.

La technologie peut être combinée avec des systèmes de fichiers comme Lustre, Spectrum Scale ou HDFS et peut aussi être déployée en parallèle d'applications Big Data, de clusters Docker ou d'un SGBD comme Oracle.

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS



STORONE

Fondé en 2011, StorOne est une start-up de stockage créée par Gal Naor, l'ex-fondateur de Storwize (le pionnier de la compression de données racheté en 2010 par IBM pour 140 M\$) et Raz Gordon (l'architecte des puces de commutation Ethernet de Galileo Technologies, le pionnier des SOC de commutation Ethernet, racheté pour 2,7 Md\$ en 2000 par Marvell).

StorOne est financé par des investisseurs institutionnels et privés (dont les fondateurs de la firme). La firme compte parmi ses soutiens financiers l'ex-président et cofondateur de Sun, Ed Zander, ainsi que l'ex-CEO de Symantec, aujourd'hui président de Microsoft, John Thompson. La firme aurait levé environ 30 M\$ et ambitionne de rester durablement privée. Ses créateurs contrôlent une majorité du capital.

STORONE A DÉVELOPPÉ un nouveau système d'exploitation de stockage, baptisé TRU (pour Total Resource Utilization), conçu pour motoriser des baies de stockage à base d'architectures x86 banalisées. Selon ses concepteurs, l'objectif de TRU est d'éliminer les goulots d'étranglement habituels des OS de stockage, afin de tirer parti au

maximum des ressources CPU, mémoire et stockage des architectures matérielles modernes.

UN OS DE STOCKAGE MODERNE

TRU est conçu pour fonctionner aussi bien avec des disques Flash (NVMe ou non) qu'avec des disques durs. « Il est incompréhensible qu'une baie de stockage Flash facture une opération par seconde 50 cents ou 1 \$, alors qu'aujourd'hui un simple SSD peut délivrer 250 000 IOPS pour moins de 500 \$ », explique Naor. Selon lui, TRU garantit qu'un serveur équipé de SSD délivrera en pratique le niveau d'IOPS théorique des disques installés dans la machine (pour peu que le CPU et le réseau suivent).

Dans la pratique, le système d'exploitation de StorOne s'installe sur tout type de serveurs (physique ou virtuel) et supporte de multiples configurations (monoserveurs, multiserveurs, avec ou sans disques flash, etc.) et permet de délivrer l'ensemble des services d'une baie unifiée moderne.

Le logiciel supporte tous les protocoles d'accès du marché, Fibre Channel et iSCSI pour les accès en mode blocs, NFS, et SMB pour les accès en mode fichiers et S3 pour le mode objet. Ces services sont déployés au-dessus d'une couche de stockage unifiée, qui prend en charge la distribution (via un mécanisme d'[erasure coding](#) à parité variable) sur l'ensemble des disques disponibles sur un système.

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Selon la firme, les données sont distribuées sur l'ensemble des disques par un mécanisme de code à effacement de type $n+k$. Les volumes sont quant à eux des conteneurs virtuels au-dessus de ce mécanisme et ne sont donc pas liés à des disques physiques spécifiques. Cela permet d'optimiser au maximum l'usage de la capacité disponible.

StorOne indique que la plate-forme assure la disponibilité et la protection des données sans impact perceptible sur la performance et qu'il est possible de gérer une infinité de [snapshots](#) par volume — avec support de VSS et des groupes de cohérence. Les fonctions de [réplication](#) entre baies StorOne, de type quasi-CDP, s'appuient d'ailleurs sur cette capacité.

STORONE REVENDIQUE DES PERFORMANCES ÉLEVÉES À UN PRIX TRÈS FAIBLE

D'après Naor, un serveur bi-contrôleur (2 nœuds Xeon E5v4 connectés à un JBOD SAS) équipé de 4 SSD et motorisé par TRU est capable de délivrer un niveau de performance similaire à celui offert par une baie de stockage moderne de milieu de gamme. La même machine avec 4 disques NVMe délivre une performance similaire à celle d'une baie haut de gamme.

Selon la firme, TRU déployé sur une configuration typique pourra délivrer jusqu'à 4 Millions d'IOPS, 18 Go/s de bande passante et gérer jusqu'à 15 Po de capacité avec un système à base de disques NVMe. Selon la firme, il est

possible de concevoir des architectures redondantes et tolérantes aux pannes avec plusieurs contrôleurs motorisés par StorOne. L'administration de TRU s'effectue via une interface HTML très conviviale, mais aussi via une interface en ligne de commande ou une API Rest.

Côté tarifaire, StorOne met en avant une politique très agressive adaptée selon la nature des usages des clients. Ainsi, pour une configuration optimisée pour la capacité, le prix du logiciel sera d'environ 6 \$ par To et par an pour une configuration de l'ordre du Pétaoctet (soit une licence annuelle de 6000 \$) et il pourra descendre aux environs de 2 \$ par To pour des configurations de plusieurs pétaoctets. Pour une configuration optimisée pour la performance, le prix annoncé est de l'ordre de 10 \$ par To pour une configuration de 150 To. Ce prix inclut l'ensemble des protocoles de stockage ainsi que l'ensemble des services de stockage (thin provisioning, protection de données, snapshot, réplication, etc.).



DRIVESCALE

L'avènement des architectures [software defined](#) pose avec une acuité nouvelle la question de la flexibilité et de la granularité de configuration des serveurs que ce soit pour des applications traditionnelles ou pour de nouveaux



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

workloads non virtualisés (Hadoop, bases de données distribuées, applications conteneurisées, etc.).

Les serveurs rack modernes n'offrent qu'une flexibilité limitée en matière de stockage et il est délicat de modifier leur configuration une fois qu'ils sont installés, car le simple ajout ou retrait d'un disque suppose une intervention manuelle.

Pour contourner ce problème, DriveScale a imaginé une architecture matérielle originale, qui permet de découpler le stockage des serveurs. La technologie « Software Composable Infrastructure » de la firme permet, à l'échelle d'un rack ou plus, d'allouer dynamiquement des disques aux serveurs depuis un pool de tiroirs JBOD SAS. Un administrateur peut ainsi reconfigurer à la volée des serveurs en fonction des workloads à faire fonctionner.

DriveScale a été fondée par des anciens de Sun et de Cisco (dont l'ex-chief Scientist de Nuova Systems, qui a créé pour Cisco la gamme UCS et les Nexus). La firme a levé 15 M\$ et compte parmi ses administrateurs et conseillers des vétérans de l'industrie. Parmi eux figurent Carl Ledbetter (ex AT&T, Control Data, IBM et Sun), Scott McNealy (co-fondateur et ex CEO et président de Sun), Amr Awadallah (CTO de Cloudera) ou Whitfield Diffie (l'un des pères de la [cryptographie](#) moderne, récompensé par un prix Turing en 2015 pour l'algorithme Diffie Hellman).

UNE SOLUTION D'ARCHITECTURE COMPOSABLE SIMPLE

Cette équipe familière des environnements de datacenter a opté pour une approche technologique simple : pour assurer la connexion physique entre les serveurs et les tiroirs de disques JBOD en mode SAS, la firme a développé ce qu'elle appelle le DriveScale Adapter. Cette appliance Rack 1U agit en fait comme un pont 10G Ethernet/SAS et permet d'attacher n'importe quel disque inséré dans un JBOD à l'un des serveurs du cluster comme s'il était une cible iSCSI.

Chaque adaptateur dispose de 8 ports 10G Ethernet et de la bande passante pour gérer environ 80 disques durs avec des performances similaires à celles de disques en attachement direct. Il est possible d'empiler autant d'adaptateurs que nécessaire au sein d'un rack et même au-delà du rack. La seule limite dans une configuration DriveScale est l'aptitude du réseau à délivrer une bande passante non bloquée de 10G entre serveurs et adaptateurs. L'autre pièce du puzzle de DriveScale est logicielle. La société fournit une série d'agents à installer sur les serveurs du cluster (sous Linux) ainsi qu'un composant d'administration, le DriveScale Management Server (DMS), qui permet de piloter l'ensemble du système.

Le DMS est un composant distribué mis en œuvre sur plusieurs serveurs, physiques ou virtuels, afin d'assurer



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

une tolérance aux pannes. Il permet de collecter les informations remontées par les agents et les « DriveScale Adapters » afin de disposer d'une vue complète des ressources disponibles dans le cluster. Son interface graphique permet à l'administrateur de composer à la volée des « serveurs logiques » à partir des serveurs physiques et des ressources disques disponibles. Selon la société, il ne faut que quelques minutes pour composer à la volée un cluster complet de serveurs prêts à l'emploi.

DRIVESCALE VISE LE MARCHÉ DES APPLICATIONS DISTRIBUÉES

Comme l'explique Tom Lyon, cofondateur et Chief Scientist de la firme, la technologie de DriveScale a d'abord été mise en œuvre dans de grands clusters Hadoop, avant que le constructeur ne valide sa solution pour les grandes bases MPP comme Greenplum, Vertica, Teradata, puis sur les environnements conteneurisés. Le DMS s'interface ainsi avec Kubernetes ou Docker Swarm (l'orchestration du stockage se faisant via le pilote de volume Docker de la firme). DriveScale fournit aussi une API Restful permettant d'interfacer le DMS avec des outils tiers.

DriveScale a certifié sa solution avec les serveurs principaux serveurs rack de Cisco, Dell EMC et HPE ainsi qu'avec les machines de fournisseurs en marque blanche comme Foxconn Quanta et SuperMicro. Elle a également noué un accord avec Dell EMC pour qu'il revende à son catalogue ses appliances DriveScale Adapter au niveau mondial. A titre de comparaison, l'approche « Composable Infrastructure » d'HPE se limite pour l'instant à un ou plusieurs chassis lames de la marque et ne sait provisionner que des disques contenus dans les lames de stockage du châssis.

Selon DriveScale, sa technologie est aussi théoriquement applicable à des serveurs lames, des équipements souvent pauvres en ressources disques. Notons, au passage, que les « Adapters » ne sont qu'une solution temporaire. Avec la généralisation des technologies NVMe et NVMe over fabrics, la société estime qu'elle pourra totalement se passer d'appliances et allouer directement des ressources NVMe over fabrics aux serveurs. ■

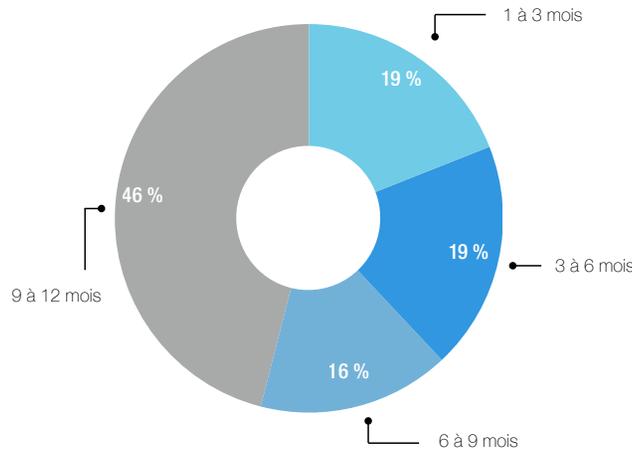
CHRISTOPHE BARDY

LES ACHATS DE BAIES SAN EN CHIFFRES

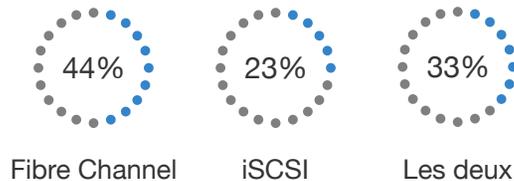
- HOME
- BLEU, BLANC, ROUGE...
- LES STARTS-UPS FRANÇAISES DU STOCKAGE ET DE LA PROTECTION DE DONNÉES
- LES OFFRES DE STOCKAGE EN CLOUD DES ACTEURS FRANÇAIS
- LES 8 START-UPS US DU STOCKAGE À SUIVRE
- INFOGRAPHIE : LES ACHATS DE BAIES SAN EN CHIFFRES
- QUEL STOCKAGE POUR HADOOP ?
- S3 : UNE OPTION DE STOCKAGE DE PLUS EN PLUS UTILISÉE POUR HADOOP
- LE STOCKAGE EST-IL TOUJOURS IMPORTANT ?
- À PROPOS



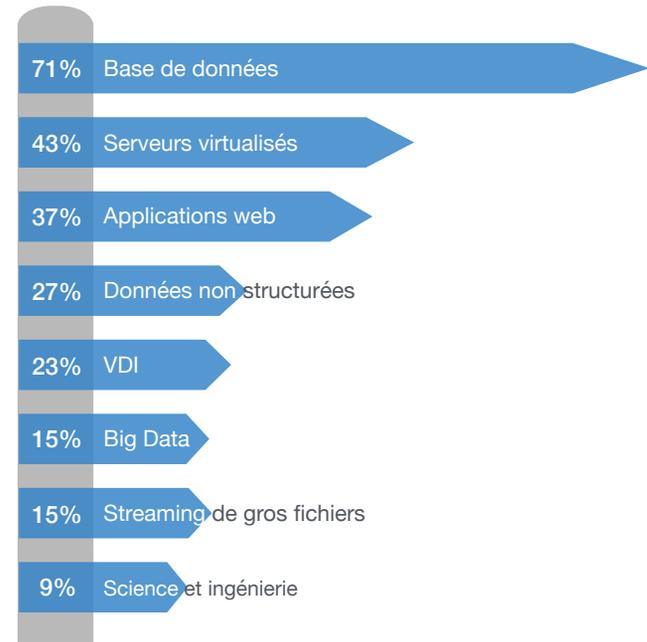
Quand prévoyez-vous d'acquérir votre prochaine baie de stockage SAN ?



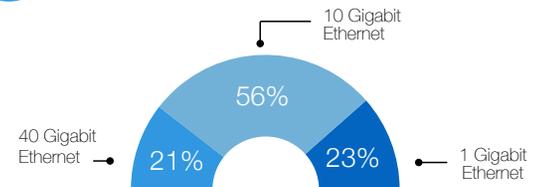
Quel protocole utiliserez-vous sur votre prochaine baie SAN ?



Quelles applications prévoyez-vous de déployer sur votre nouvelle baie SAN ?



Quel débit Ethernet utiliserez-vous sur votre prochaine baie SAN ?



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

QUEL STOCKAGE POUR HADOOP ?

HDFS, le système de gestion de fichiers distribué d'Hadoop, a considérablement évolué depuis les débuts du framework analytique. Ce qui n'empêche pas certains de lui chercher des alternatives.

Par Christophe Bardy

AU CŒUR DU FRAMEWORK OPEN SOURCE se trouve avant tout un système de fichiers en cluster, baptisé [HDFS \(Hadoop Distributed File System\)](#). HDFS a été conçu pour stocker de très gros volumes de données sur un grand nombre de machines équipées de disques durs banalisés.

À l'origine, HDFS a été conçu dans une logique de type « write once, read many », bien adaptée à la logique des traitements en mode batch de MapReduce, le framework initial d'analyse de données d'Hadoop.

S'appuyant sur le stockage local de chacun des nœuds, les premières moutures d'HDFS assuraient la sécurité des données en répliquant l'ensemble des données écrites sur le cluster. Par défaut, chaque donnée était écrite sur



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

trois nœuds différents (dont une copie sur le nœud local). Cette approche n'était à l'évidence ni la plus élégante, ni la plus efficace en matière de redondance – d'autant que la première mouture de HDFS n'opérait qu'avec un « namenode » unique, véritable faille dans l'architecture de gestion des métadonnées du cluster. Mais étant donné que l'on s'appuyait souvent sur des disques durs SATA économiques, ce dispositif permettait quand même d'offrir une solution de stockage économique par rapport aux baies de stockage traditionnelles.

Autre problème, les premières moutures d'HDFS étaient essentiellement optimisées pour maximiser les débits de données et non pas pour traiter des opérations transactionnelles aléatoires. Le fruit, là encore, de la nature batch de [MapReduce](#).

Progressivement, HDFS a donc évolué pour tenter de corriger ses limitations initiales. Avec la version 2.0 d'Hadoop, la principale faiblesse d'HDFS a été levée : la fonction HDFS High Availability a ainsi permis de répliquer le « name node » en mode actif/passif, protégeant ainsi le système de fichiers contre une éventuelle défaillance d'un des namenodes. Cela offre une tolérance aux pannes.

Plus récemment, la version 3.0 d'HDFS a apporté l'intégration native des codes à effacement. L'[erasure coding](#) (issue d'une contribution d'Intel et de Cloudera) permet

à HDFS de stocker les données plus efficacement sur le cluster et d'améliorer la récupération et la fiabilité des données. Surtout, avec l'erasure coding, Hadoop n'a plus à stocker 3 répliques complètes des données sur le cluster. Techniquement, chaque dossier peut être associé à une règle (policy) particulière qui déterminera la répartition des blocs stockés. Les fichiers créés dans le dossier hériteront alors de cette règle. Selon la documentation d'Hadoop, « comparé à la triple réplification, la mise en place de l'erasure coding permet d'économiser 50 % de capacités de stockage tout en améliorant la tolérance aux pannes ». Ces gains se traduisent bien évidemment par une réduction globale des coûts de stockage.

PASSER OUTRE LES LIMITATIONS DE HDFS

Quelles que soient les améliorations apportées à HDFS, un autre « défaut » du système de fichiers distribué est qu'il n'est pas conforme au standard POSIX. Certaines fonctionnalités et commandes familières sur un système de fichiers traditionnel ne sont donc pas disponibles. C'est ce qui a amené certains acteurs à tenter de trouver une solution de substitution à HDFS. MapR, l'un des pionniers d'Hadoop, a ainsi développé son propre système de gestion de fichiers distribué qui règle le problème de fragilité lié aux « name nodes » d'HDFS (il distribue les informations de métadonnées sur les nœuds de données). Celui-ci ajoute aussi des fonctions avancées comme les snapshots, la réplification ou le clonage. Ce système de

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

fichiers a récemment évolué pour gagner des capacités multi-datacenter, et multi-cloud (Azure, AWS) et, surtout, des capacités de tiering sophistiquées. [Rebaptisé MapR-XD, il a aussi gagné le support du protocole S3](#) ; ce qui permet à MapR d'inclure le stockage objet Cloud d'AWS à son système de stockage et à sa politique de tiering.

Plusieurs constructeurs de baies de stockage comme Dell EMC, HPE ou IBM ont aussi développé des couches de compatibilité HDFS pour certaines de leurs baies de stockage. Leur objectif est de remplacer le modèle de stockage décentralisé de HDFS par un modèle centralisé de type SAN ou NAS. Leur argument est double : le stockage centralisé est plus efficace, plus optimisé et plus simple à gérer que le stockage HDFS. Surtout, il permet d'alléger les nœuds des fonctions de stockage pour leur permettre de se concentrer aux seuls calculs analytiques. Pour les entreprises, on peut également souligner d'autres

bénéfices. Tout d'abord, elles sont déjà familières des modèles de stockage SAN et NAS. Ensuite, elles sont habituées aux services de stockage riches offerts par les baies (snapshots, réplication, etc.). Enfin, le fait de s'appuyer sur les baies qui stockent déjà leurs données primaires et secondaires évite d'avoir à transférer des données sur HDFS. Cela permet aussi de s'appuyer sur des protocoles standards comme NFS ou SMB pour verser des données additionnelles dans le cluster Hadoop.

Notons pour terminer que le stockage traditionnel n'est pas la seule alternative au stockage HDFS. De plus en plus, le stockage objet émerge aussi comme une alternative au stockage Hadoop traditionnel grâce notamment à la montée en puissance du client s3a (voir l'article suivant). ■

CHRISTOPHE BARDY

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

S3 : UNE OPTION DE STOCKAGE DE PLUS EN PLUS UTILISÉE POUR HADOOP

Au cours des dernières années, le support du protocole S3 par Hadoop s'est considérablement enrichi. Au point que de nombreux utilisateurs effectuent aujourd'hui directement des requêtes sur des données stockées dans des systèmes de stockage objet, là où auparavant, ils les auraient importées dans HDFS.

Par Christophe Bardy

DEPUIS SA CRÉATION, Hadoop s'est imposé comme le framework de référence pour l'analyse de données en volume. Le problème est qu'avant d'analyser des données, il a longtemps été nécessaire de les charger dans le cluster [Hadoop](#) sur le système de fichiers HDFS. Or, ce dernier n'est ni le moyen le plus efficace, ni le plus simple, ni le plus économique pour stocker de grands volumes de données.

Le problème est que pour alimenter les algorithmes de plus en plus sophistiqués élaborés par les datascientists, il faut de plus en plus de données. Et ces données résident



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

de plus en plus dans des systèmes de stockage objet : ils ont en effet le double mérite d'offrir une performance satisfaisante tout en ayant un très faible coût au gigaoctet. Ils sont aussi le support de stockage de choix d'une large partie des applications web de nouvelle génération. Très vite, l'idée est donc apparue [d'ajouter le support du protocole S3 à Hadoop](#) afin de permettre au framework d'accéder nativement à des données stockées dans un stockage compatible S3. Ce support a le net avantage de permettre à Hadoop d'analyser ces données sans avoir à les précharger préalablement sur les nœuds HDFS du cluster.

UN SUPPORT NATIF DE S3 DANS HADOOP

Le support de S3 dans Hadoop s'est régulièrement enrichi au cours des dix dernières années. Le premier support natif du protocole de [stockage objet](#) dans le framework est apparu en 2008 avec l'inclusion du client natif S3 (ou s3n, pour S3 Native Filesystem). L'intégration de s3n a permis à Hadoop de lire et d'écrire nativement des données dans un magasin S3 existant et donc de manipuler directement des données stockées dans un ou plusieurs « buckets », sans avoir nécessairement à les importer préalablement dans HDFS. Cette première implémentation souffrait toutefois de problèmes de performances et ne supportait pas certaines fonctions essentielles du protocole S3.

En 2014, la version 2.7 d'Hadoop a donc officiellement remplacé le client s3n par un client de nouvelle

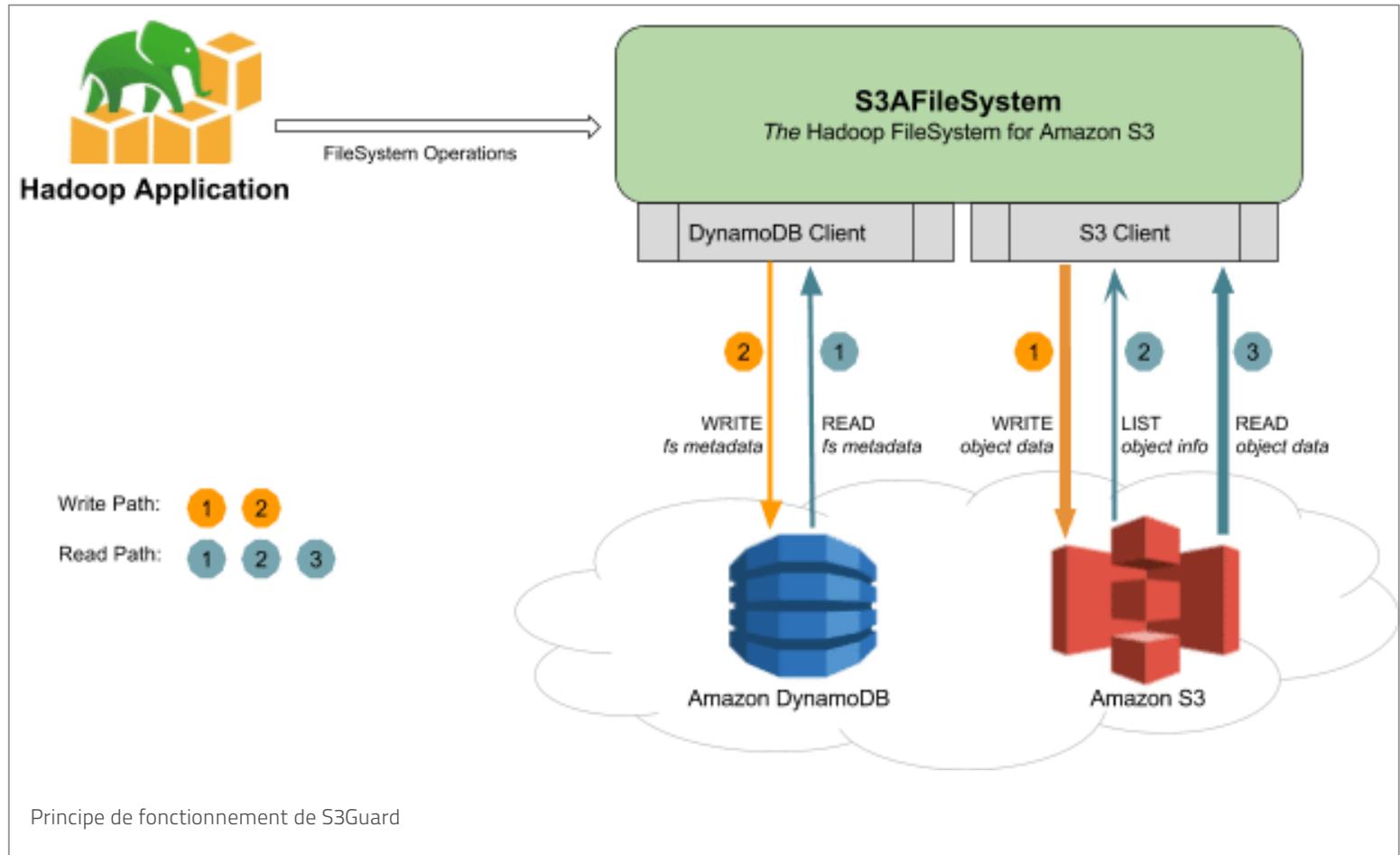
génération baptisé s3a. Ce nouveau client a apporté de multiples améliorations, dont celle portant sur des mécanismes de gestion des identités et des accès de S3. Il a également singulièrement amélioré les performances (notamment avec des outils comme Hive ou Spark).

L'arrivée de s3a n'a toutefois pas éliminé tous les problèmes que pose l'utilisation du protocole S3 avec Hadoop. En particulier, l'un des problèmes majeurs à traiter est le fait que contrairement à la plupart des stockages objet locaux, le service AWS S3 n'est qu'éventuellement cohérent. On entend par là qu'après une opération d'écriture, de mise à jour ou d'effacement, il est parfois nécessaire qu'un certain délai s'écoule avant que les modifications soient répercutées dans le système. Ainsi, en listant le contenu d'un répertoire, il est possible que S3 retourne un état qui ne soit pas cohérent avec les dernières opérations effectuées. Cela peut s'avérer catastrophique si l'on effectue une requête MapReduce, Hive ou Spark.

S3GUARD : UN MÉCANISME POUR GARANTIR LA COHÉRENCE DES DONNÉES AVEC AWS S3

Les développeurs ont donc travaillé au développement de S3Guard dont l'objectif est de se protéger contre les éventuelles incohérences d'AWS S3 en maintenant un journal des opérations d'écritures effectuées sur le service. Ce journal est conservé sous forme de métadonnées dans une base DynamoDB et permet lors d'une lecture

S3, LE FUTUR DU STOCKAGE POUR HADOOP



de s'assurer de la cohérence des informations retournées. Les métadonnées stockées dans [DynamoDB](#) pouvant être utilisées comme un cache, la mise en œuvre de S3Guard contribue également à accélérer certaines opérations de

lecture et à doper les performances des requêtes exécutées directement sur un magasin de stockage S3. Il est à noter qu'à ce jour S3Guard reste une fonction expérimentale du client s3a.

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS



HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

Hadoop : un support objet qui ne se limite pas à S3

AWS S3 n'est pas le seul service de stockage objet supporté par Hadoop. Au fil des années, le framework de traitement Big Data a hérité du support d'OpenStack Swift et d'Azure Blob.

En 2013, Rackspace et Mirantis ont contribué le code initial permettant le support du protocole OpenStack Swift. Deux ans plus tard, la collaboration entre Microsoft et Hortonworks a accouché de WASB, qui permet le support du stockage Azure Blob. Le support de WASB est activé par défaut dans HDInsights, l'implémentation Hadoop de Microsoft, mais aussi dans les distributions Hadoop d'HortonWorks et Cloudera - lorsqu'elles sont déployées dans des VM Azure.

En 2016, Microsoft a aussi ajouté le support d'Azure Data Lake comme cible de stockage pour Hadoop. La même année, Alibaba a également contribué le code nécessaire au support de son service de stockage objet Aliyun OSS de son service Alibaba Cloud (Aliyun).

CHRISTOPHE BARDY

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

LE STOCKAGE EST-IL TOUJOURS IMPORTANT ?

Les grandes tendances dans le monde du stockage semblent converger vers l'élimination de celui-ci en tant que silo IT.

Par Mike Matchett

AVEZ-VOUS SENTI UN CHANGEMENT dans le monde du stockage ? Peut-être avez-vous remarqué une certaine résignation parmi les vétérans de l'industrie lorsqu'ils envisagent l'avenir de la technologie du stockage de données ? Peut-être avez-vous constaté que lorsqu'il s'agit de rafraîchir des baies de stockage vieillissantes, il y a moins de différenciation entre les produits concurrents - ou pas de nouvelles technologies de stockage excitantes. Après tout, tout le monde a déjà une baie avec de la flash, n'est-ce pas ? Peut-être manque-t-il des guerres de vendeurs enflammées pour retrouver un peu d'excitation. Ou peut-être qu'un grand nombre de vos besoins essentiels en matière de stockage sont désormais satisfaits en utilisant un service en nuage, un produit sans nom ou même un logiciel libre.

Pendant de nombreuses années, il a été amusant de voir les grands fournisseurs de stockage se battre pour la bonne cause. Ils avaient l'habitude de s'aligner dans la première rangée lors de grandes expositions comme



LE STOCKAGE EST-IL TOUJOURS IMPORTANT ?

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

VMworld, se disputant le plus grand stand pour montrer leurs produits les plus récents. Mais depuis l'an passé, il semble que l'importance du stockage a reculé de quelques rangs. Bien sûr, les tendances du marché, telles que le stockage défini par logiciel ou l'[hyperconvergence](#), ont changé les règles du jeu dans le stockage. Mais, jusqu'alors, lorsque le jeu changeait, les fournisseurs de stockage compétitifs s'adaptaient. C'est peut-être plus difficile à faire maintenant que le stockage est de plus en plus embarqué, intégré, convergent et rendu plus accessible par les concurrents dans le cloud.

De nombreuses tendances récentes impliquent de se débarrasser du stockage en tant que silo informatique, ce qui soulève des questions sur l'avenir même des systèmes de stockage de données. Comment vendre du stockage avec un S majuscule si plus personne n'achète de stockage autonome ? Est-ce que nous arrivons à la fin du stockage en tant qu'industrie de première classe ? La réponse, pour faire vite, est non.

DES DONNÉES, PLEIN DE DONNÉES...

Les données chiffrées sur le marché du stockage montrent que les acteurs traditionnels du stockage continuent de souffrir du rétrécissement des marges pour leurs systèmes de stockage haut de gamme. Dans le même temps, le volume de stockage mondial continue de croître. Avec l'explosion des volumes de données due à la

mondialisation des applications et avec l'émergence de bases de données « Webscale », de l'analytique Big Data, de l'archivage en ligne et de l'internet des objets, il faut bien stocker quelque part tous ces nouveaux octets.

Il n'est tout simplement pas possible de stocker toutes ces données dans des services de stockage froid en cloud. Si les données valent la peine d'être conservées, il faut en extraire le plus de valeur possible. Et s'il s'agit de données importantes, il faut les gouverner, les protéger, les sécuriser et finalement les gérer activement. Même si le stockage s'effectue dans une infrastructure convergente, hyperconvergente ou en nuage, une grande partie de cette infrastructure contient encore des données d'entreprise importantes.

Il semblerait que l'avenir de la technologie de stockage de données verra les fournisseurs de stockage continuer à se concurrencer, se différencier et même à exciter leurs clients avec des services de gestion de données toujours plus puissants.

LE STOCKAGE EST PLUS IMPORTANT QUE JAMAIS ET IL LE SERA PLUS ENCORE DEMAIN

De nombreux services avancés sont superposables, ce qui signifie qu'ils ne dépendent pas étroitement des matrices de stockage ou des composants de stockage de niveau inférieur. Cela permet aux fournisseurs tiers de se multiplier, et bien que cela soit bon pour les consommateurs, cela complique le business des fournisseurs de stockage



LE STOCKAGE EST-IL TOUJOURS IMPORTANT ?

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

dont les activités sont basées sur des hypothèses de marges plus élevées.

Certains services de données avancés peuvent bénéficier d'une intégration plus étroite avec le stockage de bas niveau, bien sûr. Mais ces services et les fournisseurs qui les offrent peuvent trouver la meilleure opportunité de briller lorsqu'ils convergent en profondeur avec plus que de simples composants de stockage. Le problème pour les « pure-players » du stockage pur est que les produits hyperconvergés, qui intègrent calcul, réseau et d'autres domaines, ne sont pas confinés au silo de stockage et prennent des parts de marché visibles aux acteurs du stockage.

RÉGLER LE STOCKAGE SUR LE MODE AUTOMATIQUE

Quand il s'agit de l'avenir de la technologie de stockage de données, peut-être que ce qui est en question est plus l'administration du stockage que la façon dont la capacité de stockage est délivrée. L'automatisation accrue des technologies de l'information est-elle le véritable coupable ? Il est possible que le stockage d'entreprise traditionnel - et son exigence d'expertises pointues en matière d'administration - soit destiné à être supplanté par des composants entièrement automatisés, intégrés à des architectures de type cloud. Si c'est le cas, les experts du stockage doivent-ils se préparer à fermer boutique et à rentrer chez eux ?

Ne vous méprenez pas sur mes propos : l'automatisation informatique est une bonne chose. Elle est au cœur de la plupart des nouvelles initiatives IT, y compris l'orchestration, le libre-service et le DevOps. Augmenter l'automatisation est une évidence ; l'informatique devrait toujours s'efforcer d'automatiser les tâches répétitives, onéreuses et sujettes aux erreurs. De plus, l'automatisation validée par le fournisseur permet d'intégrer de nombreuses tâches de gestion du stockage à des systèmes de plus grande taille, masquant efficacement la complexité de stockage.

Il y a toutefois un rayon de lumière pour les professionnels du stockage. Lorsqu'une organisation écrit et déploie un script d'automatisation, elle crée une dette technique. À un certain moment - ou même à de nombreux moments - dans l'avenir, cette automatisation devra probablement être optimisée, améliorée, déboguée, mise au point, profilée ou évoluée.

LE MEILLEUR DES MONDES POUR LE STOCKAGE

C'est peut-être impopulaire à dire, mais je pense que les besoins en matière d'expertise de stockage sont en train d'évoluer vers un style de gestion opérationnelle dynamique étroitement lié au modèle DevOps. Nous pourrions appeler ce futur rôle OpsDev, dans lequel l'opérateur d'infrastructure, maintenant agile, programme et



LE STOCKAGE EST-IL TOUJOURS IMPORTANT ?

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIES
SAN EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

reprogramme dynamiquement une pile informatique de plus en plus définie par le logiciel, y compris les services de stockage et de gestion des données associés.

Les fonctions d'assistance et de support à distance des fabricants devenant plus dynamiques et en temps réel avec chaque version et les algorithmes d'apprentissage machine s'insinuant dans presque toutes les facettes de l'administration des systèmes informatiques, vous pouvez arguer que l'expertise en stockage autonome deviendra bientôt inutile. Mais peu importe à quel point l'administration à distance devient intelligente, un professionnel sur site qui connaît l'entreprise et ses exigences en matière de données, qui sait ce qui est acceptable pour les utilisateurs et comprend les autres exigences spécifiques, conservera toujours la responsabilité côté entreprise.

Nous avons également encore de grands défis à relever en ce qui concerne l'avenir des technologies de stockage de données. Le stockage pourrait bientôt être moins une question de choisir tous les trois ou cinq ans des baies

spécifiques avec de nouvelles fonctionnalités qu'une question d'assemblage d'un ensemble de services de stockage, aussi économiques, évolutifs et distribués, déployables à l'échelle mondiale. Et nous devons faire tout cela tout en délivrant les performances attendues, un accès instantané aux jeux de données Big Data, en satisfaisant les besoins de capacité à la demande, tout en maintenant une gouvernance d'entreprise complète et en assurant la protection et la sécurité des données.

Peut-être que la fonction ne s'appellera même pas administrateur de stockage dans quelques années, mais plutôt quelque chose comme facilitateur en chef de l'accès aux données. Qui sait ? Le fait est que ce n'est vraiment qu'une nième évolution du stockage. Quel que soit le nom que vous lui donnez, le stockage reste plus important que jamais. Et demain, ce sera encore le cas. ■

MIKE MATCHETT est analyste senior chez *Small World Big Data*.
Suivez-le sur Twitter [@smworldbigdata](https://twitter.com/smworldbigdata).

HOME

BLEU, BLANC, ROUGE...

LES STARTS-UPS
FRANÇAISES DU STOCKAGE
ET DE LA PROTECTION
DE DONNÉES

LES OFFRES DE STOCKAGE
EN CLOUD DES ACTEURS
FRANÇAIS

LES 8 START-UPS US DU
STOCKAGE À SUIVRE

INFOGRAPHIE :
LES ACHATS DE BAIE DE
STOCKAGE EN CHIFFRES

QUEL STOCKAGE POUR
HADOOP ?

S3 : UNE OPTION DE
STOCKAGE DE PLUS EN
PLUS UTILISÉE POUR
HADOOP

LE STOCKAGE
EST-IL TOUJOURS
IMPORTANT ?

À PROPOS

SEARCHSTORAGE.FR

RÉDACTEUR EN CHEF **Cyrille Chausson**

RÉDACTEUR EN CHEF ADJOINT **Christophe Bardy**

COORDINATRICE EDITORIALE **Pascale Roncin**

ÉDITEURS **Bill Crowley et Byrony Seifert**

ABONNEMENT

www.SearchStorage.fr

TechTarget - LeMagIT, 22 rue Léon Jouhaux, 75010 Paris

©2018 TechTarget Inc. Aucun des contenus de cette publication ne peut être transmis ou reproduit quelle que soit sa forme sans l'autorisation écrite de l'éditeur. Les réimpressions des publications de TechTarget sont disponibles via les services de The YGS Group.

TechTarget édite des publications pour les professionnels de l'IT et propose plus de 100 sites qui fournissent un accès rapide à un stock important d'informations, de conseils, d'analyses concernant les technologies, les produits et les processus essentiels pour vous aider dans vos fonctions. Nos événements et nos séminaires virtuels vous donnent accès à l'expertise et aux recommandations d'experts sur les problèmes et défis que vous rencontrez quotidiennement. Notre communauté en ligne "IT Knowledge Exchange" (Echange de connaissances IT) vous permet de partager vos questions et d'échanger des informations avec vos pairs et des experts du secteur.