

# Silicon Valley Tour : Un stockage toujours plus logiciel et objet.

La haute disponibilité, la gestion de gros volume de données et la reprise sur incident ne sont plus l'apanage de grandes entreprises disposant de budgets informatiques en millions de dollars. Une démocratisation possible grâce au logiciel !

## **Hedvig : du SDS plus granulaire et intelligent**

En juin 2012, Hedvig est lancée avec un capital d'amorçage de 2,5 millions de dollars (Redpoint Ventures et True Ventures) afin de finaliser son offre de Software Defined Storage. Fin mars 10 millions sont apportés par les investisseurs rejoints par Atlantic Bridge. Début juin 2015, un autre groupe d'investisseurs ajoute encore 18 millions de dollars. Soit plus de 30 millions de dollars en trois ans...

«*Nous n'avons pas cherché à être financés, les investisseurs sont venus à nous,*» s'étonne encore Avinash Lakshman, fondateur et CEO de Hedvig (après avoir coinventé Dynamo NoSQL chez Amazon Web Services et Cassandra chez Facebook).

Quel nom étrange ! Hedvig est en fait l'acronyme de: "hyperscale, elastic, distributed, virtualized, intelligent and granular". On comprend alors qu'il sera question de Software defined Storage.

«*Provisionner de l'espace de stockage peut prendre plusieurs heures, jours ou semaines. Cela n'est pas raisonnable à l'heure du cloud, ou les géants comme Google, Ebay, Netflix, Facebook... ont montré la voie d'un nouveau stockage où du logiciel associé à des serveurs standards, devenus de simples commodités, permettait d'obtenir du Software Defined Storage : élastique jusqu'à plusieurs pétaoctets de données, simple pour provisionner tout type de stockage (fichier, bloc ou objet) et flexible pour se connecter à tout serveur, puissance de calcul ou cloud,*» explique Rob Whiteley, VP Marketing de la start-up. «*Certes, le SDS est surtout accessible à ces sociétés. Désormais, en associant le logiciel Hedvig et de serveurs standard x86 ou ARM, ces technologies sont à la portée de toute entreprise.*»

La plateforme de stockage distribuée Hedvig repose sur un cluster exploitant la puissance des systèmes distribués, la simplicité avec de multiples fonctions évoluées pour concevoir un système de stockage à base de serveurs standard pour tout application, hyperviseur, container ou cloud.

L'unité exposée par Hedvig est un disque virtuel totalement personnalisable selon les besoins de l'entreprise.

Le Storage Service est moteur distribué (objet d'un brevet) faisant évoluer le stockage de façon prédictive et dynamique. Il propose toutes les caractéristiques de résilience et d'efficacité d'un cluster: équilibrage de charge -stockage et calcul; auto-tiering, séquençage intelligent des entrées-sorties; auto-réparation des nœuds sans interruption; données répliquées en local, multisites ou cloud; système de répartition large sur les nœuds pour ajuster performances et résilience... Et bien sûr les services de stockage par disque virtuel : configurations All-flash ou hybride, réplication/reprise sur incident, cache d'accélération, compression et déduplication en ligne, thin provisioning, snapshots et clones à zéro impact.

Le Hedvig Storage Proxy apporte le support de tous les protocoles de stockage pour un stockage évolutif ne nécessitant aucune modification des applications. Il s'exécute sur une VM ou un container Dockers pour supporter tous les environnements d'exécution comme VMware vSphere, Microsoft Hyper-V, KVM, OpenStack (via Cinder) et Xen. Et il gère tout type de stockage : Bloc (iSCSI), Fichier (NFS) ou objet (S3 et Swift). Enfin des API Rest ouvrent les fonctions de la plateforme aux développeurs. Une architecture intéressante qui devrait faire parler d'elle très rapidement.

## **Basho conjugue NoSQL et stockage objet à haute disponibilité**

Depuis 2009 (un an après sa création), Basho Technologies a levé 57,5 millions de dollars en sept tours. Les investisseurs misent sur les technologies de systèmes distribués qui se décline en trois produits : la base de données NoSQL Riak KV, le système de stockage Riak S2 et la Basho Data Platform regroupant les deux et offrant un service complet de base données NoSQL tout intégré. Des solutions en version open source sont déclinées en versions commerciales avec des fonctions plus évoluées.

Forte de 120 employés (Washington, Tokyo, Londres), la société affiche déjà plus de 200 clients, dont plusieurs contrats à plusieurs millions de dollars auprès de clients comme Microsoft pour Yammer, BestBuy, Comcast... S'inspirant des démarches de Google, Yahoo ou Facebook, ayant développé leurs bases de données parce que celles du marché ne convenaient pas, Basho a conçu sa base de données NoSQL pouvant exécuter des requêtes classiques (SQL...) et des traitements sur de l'information non structurée. Autres caractéristiques: évolutivité horizontale linéaire (ajout de serveurs), possibilité de localisation géographique pour rapprocher l'information du consommateur, et haute disponibilité (insensible aux pannes des serveurs, réseaux ou datacenters).

*« La reprise après incident classique (recovery) implique un arrêt du système. Il faut qu'une panne intervienne pour déclencher cette opération. Nous avons donc plutôt intégré l'élasticité (resiliency), qui permet de passer de ce mode réactif à un mode proactif. Dans ce cas, l'infrastructure permet à l'entreprise de poursuivre ses activités malgré un incident, grâce à des systèmes distribués dupliquant les informations et les tâches,»* explique Dave McCrory, directeur technique chez Basho.

La base de données Riak KV fonctionne en mode cluster évolutif et à haute disponibilité avec des données systématiquement recopiées sur trois nœuds. La base propose des API et de nombreux clients logiciels ( Java, Ruby, Python, Erlang, .Net, Node.js, PHP...).

Moins connue, et pourtant de plus en plus sollicitée, la solution de stockage objet Riak S2 (topologie en anneau) est hautement disponible, évolutive, *«et simple à utiliser pour stocker des images, des textes, des vidéos, des documents de sauvegardes de bases de données et des applications,»* précise l'éditeur. Compatible avec Amazon S3 et Openstack Swift, Riak S2 propose aussi des API dans le but d'agréger des pétaoctets sur des équipements standards avec une évolutivité quasi linéaire des performances.

*« Les logiciels NoSQL ont évolué de solutions ponctuelles vers des solutions multi-modèles. Cependant, les entreprises passent beaucoup de temps à déployer et maintenir des solutions différentes pour gérer le big data, l'analytique, l'internet des objets, etc. une complexité croissante lorsqu'il s'agit de déployer des services à travers des environnements cloud publics, privés et hybrides,»* analyse Peter Coppola, vice-président

Produit chez Basho. Au final, plusieurs clusters à superviser (Riak, Redis, Spark...) de multiples opérations manuelles de maintenance, voire de synchronisation...

Pour simplifier l'exécution et la gestion de ces piles technologiques hétérogènes, Basho propose donc sa Data Platform, reposant sur une couche de stockage avec des modules conçus ou intégrés par Basho, les services de données Riak au cœur du système (avec réplication/synchronisation, gestion de cluster -éliminant le besoin de Zookeeper, data store, message routing, logs et tracking), et des extensions de services de données développés par Basho (Aopache Sparck Add-on, Redis Add-on...) ou par d'autres.

Avec cette plateforme, Basho automatise une grande partie des tâches des multiples modules spécialisés pour orchestrer l'ensemble des besoins en bases de données NoSQL, et plus. L'infrastructure distribuée en cluster apportant performance et haute disponibilité sur des équipements standards avec une évolutivité quasi linéaire... Autant d'arguments séduisants pour cette toute nouvelle plate-forme à suivre.

### **Catalogic optimise la gestion des copies de données**

En octobre 2013, le spécialiste mondial de la sécurité Syncsort revend son activité de protection des données à des investisseurs. Cette spin-off prend le nom de Catalogic Software en janvier 2014.

Editeur de solution de copie de données intelligente, Catalogic se positionne pour optimiser la copie de données : le software Defined Data Copy (un SD de plus), illustrant l'automatisation de ces tâches encore manuelles ou sous des environnements hétérogènes . En effet, entre les environnements de test et de développement, les sauvegardes et les répliquions, les copies applicatives... on compte souvent plusieurs dizaines de copies des données et dont une partie reste totalement inactive, sans que personne ne le sache.

*« Notre plate-forme permet de cataloguer les copies, de les administrer et d'en optimiser le nombre, sans agent à installer, »* explique Steve Kenniston, vice-président Marketing et Stratégie chez Catalogic. *« Pas de rupture technologique, pas de remplacement : tout s'installe en quelques minutes pour automatiser les environnements de test/développement, la reprise sur incident, l'alimentation des applications analytiques, etc. »*

Les solutions Catalogic fonctionnent avec les environnements NetApp (7-Mode, Clustered Ontap, Cloud Ontap), VMware, et bientôt (ECX 2 .5) IBM (Storwize, SVC, v9000 and Flash Copy Manager). *« Dès 2016, nous intégrerons aussi les environnements EMC, »* annonce Steve Kenniston.

Avec DPX, l'entreprise dispose d'une interface unique pour gérer et coordonner tous ses équipements virtuels, physiques et cloud. La solution intègre sauvegarde et restauration, reprise sur incident, sauvegarde sur bande, avec une gestion de copies de données favorisant la réduction de leur nombre, et donc de l'espace nécessaire, du temps de traitement et du nombre d'objets à maintenir.

ECX, solution phare de Catalogic, est une plateforme gérant le cycle de vie complet des copies de données. Après identification des fichiers et applications de l'entreprise, elle offre une gestion automatisée de la mise en place et du suivi des politiques de création des copies (Snapshots, SnapMirror, SnapVault...), l'orchestration permettant de simplifier et automatiser l'exploiter les données en mode critique (reprise d'activité, tests/développements, analytique...) et le reporting

(suivi en temps réel de l'état des politiques de copies, respect des SLA, recherche pour optimisation et ajustement, etc.).

**A lire aussi :**

[Silicon Valley Tour : 100% ou hybride, le flash devient incontournable](#)

[Silicon Valley Tour : Le Software Defined Network toujours hyperactif](#)

**crédit photo : jijomathaidesigners-Shutterstock**