

EMC DSSD D5: le stockage à la vitesse de la mémoire

Après [l'annonce du rachat de DSSD](#) lors du EMC World en mai 2014, et [une première présentation furtive](#) (mais détaillée) lors du même évènement en mai 2015, EMC annonce la disponibilité de son équipement de stockage All-Flash à très hautes performances DSSD D5. Ce dernier affiche des performances impressionnantes: 144 To de capacité, 100 Go/s de bande passante, 10 millions d'I/O et une latence de 100 microsecondes. Ce qui le positionne plutôt face à des équipements de type supercomputing. Il y a quelques semaines, [nous avons rencontré la start-up Mangstor](#) dont la solution ressemble furieusement au DSSD. Un créneau qui verra sans doute arriver d'autres constructeurs.

La nouvelle voie du rack-scale Flash

Directement connecté à un ordinateur ou à serveur, le D5 n'est pas vraiment une baie ni un équipement réseau. EMC a donc opté une nouvelle catégorie: le Rack-Scale Flash qui semble bien aller à ce rack 5U.

Pour obtenir des performances extrêmes en stockage, il convient de jouer sur plusieurs paramètres principaux : une capacité conséquente et performante, une bande passante maximale, de nombreuses entrées/sorties performantes, et un temps de latence minimal. Imaginez que la mémoire des serveurs échange quasi directement avec le support de stockage. Non pas avec la baie de stockage, mais bien avec le support lui-même, et un support de type Flash. Pour faire bref, c'est plus ou moins ce que se propose de faire le DSSD D5.

Une densité élevée sur rack 5U

Cet équipement au format 5U (pouvant donc s'intégrer simplement dans une armoire avec des serveurs racks) est composé de trois emplacements principaux sur la face avant (voir illustration).

Au centre, il peut recevoir jusqu'à 36 modules flash de 2 ou 4 To chacun, installés sur un bus PCIe et intégrant la technologie NVMe qui permet une connexion directe entre PCIe et contrôleur flash, de réduire la latence en parallélisant les traitements via plus d'entrées-sorties (I/O) et le multipathing (couplage de plusieurs ports PCIe).

Encadrant le module central, deux tiroirs de contrôle intègrent des contrôleurs redondants (dual), et assurent la séparation entre le plan de contrôle et le plan de données afin d'optimiser la bande passante et la fiabilité. Nous verrons plus loin comment l'API Flood Direct Memory du DSSD réduit la pile I/O, donc la latence. Enfin, ce contrôleur intègre le mécanisme Cubic RAID multidimensionnel d'EMC, dont l'équipementier assure qu'il procurerait une protection des données avec un overhead limité à 17% (préinstallé en configuration Flash complète ou demie: 18 ou 36 modules).

Connexion éclair pour passer sous 100 microsecondes

En face arrière, 96 ports PCIe Gen3 répartis sur 8 cartes (4 sur chaque tiroir) offrent une connexion redondante vers 48 serveurs. Chacun des deux tiroirs (ou modules, voir illustration) assure une gestion redondante des I/O ainsi que la connexion directe aux modules de contrôle et aux modules flash (une carte PCIe Gen3-8 lignes est fournie par DSSD pour les serveurs).

Au total, le D5 affiche propose donc une bande passante de 100 Go/s, 10 millions d'entrée/sorties et une latence moyenne de 100 microsecondes pour une capacité globale de 144 To (pour environ 100 To utilisables selon EMC), le tout dans un rack 5U. Excusez du peu !

Bien entendu, pas de déduplication, de compression, ou autres snapshots. « *L'objectif vise précisément la performance brute,* » répète EMC. « *Si l'entreprise le souhaite, elle peut ajouter ce type de service en dehors du D5. Mais, ici, ce n'est pas le but recherché. Pour cela, nous proposons Xtremlo et Vmax All Flash.* »

Quel type d'objets stocke le D5 ?

Le D5 stocke des objets (sur des fragments de 4K) qui peuvent être aussi bien des blocs ou des fichiers que des paires clé/valeur (key/value), mais aussi des volumes ou des répertoires (directories). Côté organisation, des volumes contiennent des objets. Les répertoires permettent de hiérarchiser les objets dans un volume.

Les métadonnées des objets sont stockées avec les objets, et ne nécessitent donc pas d'I/O supplémentaire. L'équipement autorise l'accès des objets à tous les clients, charge à l'application appelante de gérer les accès concurrents, etc. Le management du D5 s'effectue par ligne de commande ou interface Web.

Une stack forcément optimisée

Pour parvenir à ces performances optimales, la pile logicielle joue aussi un rôle primordial. « *L'installation de Flash dans les baies de stockage permet d'atteindre des performances de l'ordre de 500 microsecondes. Ce qui représente déjà un cap important,* » explique Mike Shapiro, cofondateur de DSSD et VP Software chez EMC/DSSD. « *Cependant, pour atteindre 10 microsecondes, il fallait réduire la pile logicielle, ce que nous avons réalisé avec DSSD.* » Comme l'illustre le schéma, les parties réseau deviennent inutiles, ainsi que les divers drivers intermédiaires. Et la Fabric PCIe interne fonctionnant à très grande vitesse... Ceci explique beaucoup cela.

Comment lire ou écrire des données sur le D5 ?

EMC a baptisé les logiciels DSSD du nom de Flood (flux ou flot). Avec ce logiciel tous les objets stockés sur le D5 sont accessibles simultanément (pas de notion de modes) pour tout client.

Une fois la carte DSSD cliente installée dans l'ordinateur client (ainsi que les logiciels) et connectée

au D5, les objets sont accessibles ou peuvent être stockés soit via le DSSD Block Driver (intégré à la carte), soit via l'API Flood Direct Memory. Via l'API l'appel de données est effectué et la carte interagit directement avec le module de contrôle du D5 pour envoyer l'information, sans autre intervention logicielle. Pour un appel via le Block Driver, la requête est envoyée à l'OS de la carte, puis le circuit décrit ci-dessus s'effectue de même. Un minimalisme qui explique les performances. Bien entendu, Flood propose aussi une API permettant à des applications d'accéder nativement au D5 (la libflood C-library).

Enfin, l'autre moyen d'accès passe par des plug-ins. Ainsi, DSSD a conçu un plug-in pour Hadoop (assez simple, puisqu'un objet stocké peut être de type clé/valeur). « Le premier d'une longue série,» annonce l'équipe DSSD.

Plus de précision sur les caractéristiques techniques détaillées, visitez [cette page](#).

Le DSSD D5 est proposé en quatre déclinaisons : D5 1-2T18 avec 18 modules flash de 2To, D5 1-2T36 avec 36 modules flash de 2To, D5 1-4T18 avec 18 modules flash de 4To et D5 1-4T36 avec 36 modules flash de 4To.

A lire aussi :

[EMC World 2015 – XtremIO 4.0 et DSSD D5 : EMC flash tous azimuts](#)

[Stockage : EMC rhabille ses VMax de Flash du sol au plafond](#)

Crédit Photo : EMC