

# Michel Alliel, HDS: « les objets connectés nécessitent un stockage en mode Scale Out»

L'internet de objets va dans aucun doute générer d'importants flux et volumes de données uniatirement peu volumineuse. Peut-on réellement distinguer différents types d'objets?

Quelles infrastructures et technologies de stockage sont les plus adaptées?

Entretien avec Michel Alliel, directeur marketing Produits et Solutions chez Hitachi Data Systems.

## **Comment appréhendez-vous le terme galvaudé d'Internet des objets? Quel lien avec les objets connectés ? Et quel rapport avec Big data?**

Chez Hitachi Data Systems, nous distinguons deux familles d'objets connectés. Les premiers ne disposent pas complètement de leur propre personnalité, mais prennent celle des personnes qui les utilisent. Ils jouent le rôle de sources de données supplémentaires personnelles d'un utilisateur. En outre, elles sont rarement autonomes et fonctionnent souvent en complément d'un smartphone, d'une tablette ou d'un service cloud. Parmi ces objets : montres, bracelets, etc.

La seconde catégorie regroupe plutôt "les machines connectées à Internet" : voitures, trains, feux de signalisation, appareils médicaux, équipements informatiques...Par exemple, depuis les Jeux olympiques de Londres, Hitachi gère les données des trains des liaisons à grande vitesse britanniques générant plus d'un téraoctet de données par jour via des capteurs liés à leur fonctionnement. Plus globalement, il s'agit de superviser ces appareils en continu dans le cadre du maintien en conditions opérationnelles, incluant de l'analyse prédictive.

L'internet des objets représente la troisième vague en termes de volume du Big Data. Ce dernier est apparu pour relever le défi des gros volumes d'informations structurées des bases de données et applications. Puis, la seconde vague a répondu à la nécessité de maîtriser les informations peu ou semi-structurées du Web (logs, parcours, etc.), des réseaux sociaux ou des documents bureautiques.

## **En fait, vous distinguez les objets connectés plutôt grand public et les machines connectées B2B...**

Hitachi Data Systems s'adresse essentiellement aux B2B (ou au B2B2C) et aux secteurs verticaux comme l'énergie, le médical, les transports, etc. Par ailleurs, la plupart des objets connectés destinés aux particuliers ont pour "personnalité" celles du smartphone ou de la box internet auquel ils sont reliés. Ils ne disposent pas de leur propre personnalité, et se contentent de servir de source de donnée pour ces entités. En revanche, si les usages deviennent importants, et surtout différenciés, ils pourront devenir une source intéressante pour des applications B2B.

Depuis plusieurs années, tous les équipements que nous vendons à nos clients sont connectés et envoient en continu des données sur les machines et les environnements à nos centres de support via internet. Ces informations sont alors traitées à des fins de supervision et de maintenance par

des applications SAP et Hadoop, entre autres.

**Pourtant, on a déjà constaté des problèmes d'accès, donc d'infrastructure sur les services liés aux objets connectés.**

Si l'on considère l'internet des objets individuels (objets connectés grand public), il existe autant de standards que d'objets. En outre, la problématique ne repose pas réellement sur les volumes de données, mais plutôt sur l'applet ou la web application censée gérer ces informations. Une application est souvent dédiée à chaque type d'objet. Résultat, les serveurs sont rapidement surchargés par un trop grand nombre d'utilisateurs simultanés. C'est pourquoi certains de ces services en ligne sont souvent indisponibles le week-end lors de forts usages personnels (comme les objets connectés d'assistance aux sports).

Et le problème n'est pas lié au back-office (comme le stockage), mais intervient plutôt sur le front-end (Webapps). D'où la nécessité de nouvelles approches applicatives massivement parallèles, reposant sur une infrastructure de type IaaS (avec puissance de calcul, réseau et stockage flexibles) avec un stockage distribué, proche du traitement applicatif. Pour y parvenir, l'accès à des fermes de serveurs s'impose, avec des capacités de grille de calcul extensibles à plusieurs centaines de nœuds sur des architectures de type OpenStack. Ainsi, notre solution HCP (Hitachi Content Platform) regroupe des nœuds pouvant stocker jusqu'à 50 To par nœud, regroupés par blocs de 120 nœuds pour une capacité totale plus de 4 pétaoctets utile. Et ses modules sont empilables.

**Certains mettent en avant la nécessité d'adapter le bon mode de stockage. Fichier, bloc ou objet: Comment s'y retrouver?**

Concernant les types de stockage, on retrouve traditionnellement le stockage en mode bloc pour les données très structurées (bases de données ou applications d'entreprise), et le stockage objet pour les données non structurées. Les informations provenant des objets connectés (ou Internet des objets) représentent certes un volume considérable. Néanmoins, il s'agit de très nombreuses informations de très petite taille unitaire. Dans ce cas, le stockage le plus efficace reste le système de fichiers en mode parallèle (Scale Out). En effet, le mode bloc n'apporte rien, et le stockage objet engendrerait plus de métadonnées à traiter que de données utiles.

Dans l'exemple britannique précité, chaque train envoie un flux de données qui est capté par une plateforme Web, décortiqué, analysé et indexé en temps réel. Les données sont produites par une application intégrée à l'objet, puis reçue par une autre sur un serveur. L'un des principaux challenges consiste à pouvoir déployer à la volée des centaines de milliers d'applications connectées chacune à un objet ou à un stream de données. Bien entendu, gérer plus de 700 To de données par jour pose également le problème du stockage de la gestion du cycle de vie des données. Pour disposer efficacement des informations en temps réel, un stockage proche de l'application s'impose, sans faire intervenir de baie ni de serveur de stockage dédié. L'infrastructure est ainsi composée de nœuds exécutant à la fois une ou plusieurs applications avec le stockage réparti sur chaque nœud du même cluster.

Et aujourd'hui, il est tout à fait possible de faire tenir tout cela dans un équipement 2U avec 50 To d'information, la mémoire, la puissance de calcul et les entrées-sorties nécessaires.

**A lire aussi :**

[HDS met la main sur le Français Oxya, spécialiste de l'infogérance SAP](#)