

Auchan ajoute du Big Data à la sauce Splunk à son panier e-Commerce

Le Big Data c'est bien, le déployer dans des projets concrets, c'est mieux. Et l'innovation emprunte parfois des voies inattendues. Ainsi, une solution conçue pour analyser les données machines peut devenir la pierre angulaire Big Data pour l'entreprise et toutes ses applications métier, comme le prouve ce projet chez e-Commerce Auchan.

Une expérience en ligne (et mixte) éprouvée

Si la mode des Drive (préparation des courses alimentaires sur Internet et récupération en voiture dans un entrepôt de proximité) fait fureur aujourd'hui, Auchan avait déjà lancé son premier site en 2000 (Leers, 59). En revanche, l'enseigne ne fait pas figure de pionnier en ouvrant auchandirect.fr en 2001. Dès 2006, le distributeur rachète le site de vente informatique en ligne grosbill.com et regroupe avec toutes ces marques, ainsi que son nouveau auchan.fr dans une nouvelle division e-Commerce Auchan.

Auchandirect.fr emploie aujourd'hui 500 employés sur quatre régions (Paris, Lyon Lille Marseille). En 2014, le site a réalisé 116 millions de chiffre d'affaires pour 745 000 commandes parmi ses 7000 références, sur des acheteurs intramuros et Grande Couronne des grandes villes. Quant au portail de la marque (également à vocation commerciale), il a totalisé 90 millions de chiffre d'affaires en 2014, avec ses 200 collaborateurs.

A l'instar du secteur, Auchan mène une politique multicanal, ou plus précisément "omnicanal". Cependant la stratégie du distributeur va plus loin encore. *«Notre force et notre différence viendront de la combinaison de parcours de courses pour des clients de plus en plus connectés : avant, pendant et après l'achat,»* précise Stéphane Duarte, responsable du système d'information chez e-Commerce Auchan (back-office, décisionnel et Big Data, systèmes d'échanges et urbanisation). *«e-Commerce Auchan contribue à l'ambition multicanal du Groupe afin de répondre au plus près aux nouveaux modes de consommation des clients.»*

De la supervision de l'infrastructure matérielle...

Fin 2013, les informaticiens de la division e-Commerce expriment le besoin se s'outiller pour disposer d'un système global d'alerte en cas d'erreur sur l'infrastructure. *« A l'époque nous disposions de multiples outils pour superviser les environnements hétérogènes de notre infrastructure,»* rapporte Rached Tayech, responsable Décisionnel et Big Data chez Auchan.

« De plus, les volumes de données étaient (et sont toujours) en croissance explosive,» ajoute Stéphane Duarte.

Certes, une première amélioration avait été apportée avec le déploiement de la solution QlikView générant des tableaux analytiques globaux de l'infrastructure. Néanmoins, ces tableaux étaient générés à la demande après collecte et injection des données à chaque requête. Soit des délais qui s'évaluaient d'une à trois heures. *« Cette vue globale permettait effectivement de répondre globalement, avec des questions au cas par cas,»* souligne Rached Tayech. *« Nous avons bien tenté de mutualiser les requêtes, mais leur nature ou leur propos divergents rendaient généralement l'exercice impossible. En outre, les requêtes étaient figées et en mode asynchrone.»*

Courant 2013 et 2014, la solution Splunk est alors testée pour ces besoins de supervision de l'infrastructure chez e-Commerce Auchan. *«Il nous fallait une solution de type Big Data pour gérer de gros volumes de données hétérogènes en temps réel et ne nécessitant pas des mois de déploiement,»* se souvient Stéphane Duarte. Et justement, la solution Splunk utilise un stockage de données distribué (de type HDFS), avec des traitements parallélisés, reposant sur un moteur spécifique éprouvé, mais qui n'est pas une implémentation Hadoop (mais si le produit maison Hunk assure un lien vers des environnements Hadoop).

... à une plate-forme Big Data globale IT et métiers

Au vu de l'intérêt et de l'efficacité de la solution, les informaticiens décident de se pencher sur son utilisation dans le cadre d'applications métier. Les applications ne génèrent-elles pas elles aussi des données? Et corréliser de multiples informations éparses pour décider mieux n'est-il pas le souhait de tout utilisateur et l'ambition de tout informaticien ? En outre, de nombreuses entreprises utilisatrices de Splunk ont déjà mené ce type de projet avec Splunk.

Courant 2014, l'entreprise décide donc d'utiliser Splunk "au-delà de l'informatique", et de le déployer comme plate-forme Big Data globale. Un pilote est donc réalisé pour produire des tableaux dynamiques de bord puisant les informations directement dans les données opérationnelles. Un projet qui concerne la relation client afin d'obtenir une vision 360 degrés du client (e-client en l'occurrence) en temps réel, mais aussi le marketing pour l'analyse des comportements client, dans oublier l'informatique qui verra se concrétiser le projet de supervision globale de l'infrastructure. L'ensemble étant cohérent, puisque les données applicatives reposent sur une infrastructure à même de procurer des données machine enrichissant l'information client (adresses IP, localisation géographique, nombre de tentatives d'achat...).

« En relation client et en marketing, les besoins sont couverts par des outils faits maison, suite à des projets essentiellement menés par des utilisateurs métier. Ils sont efficaces, mais montrent aujourd'hui leurs limites pour aller au-delà du fonctionnel,» constate Stéphane Duarte.

Les tableaux de bord de ce type d'outils (aussi jolis et ergonomiques soient-ils) restent tout de même complexes pour des non-initiés. *«Au premier contact, les utilisateurs s'exclament très vite "Oh la la, c'est de l'informatique!",»* reconnaît le responsable informatique. *«Cependant, après quelques informations et un petit accompagnement, il suffit de laisser Splunk entre leurs mains. Alors, c'est eux qui expriment de nombreuses demandes.»*

Une approche technologique à portées multiples

Avec Splunk, plusieurs types d'approches parallèles sont possibles. Les informaticiens accèdent directement aux données et les utilisateurs ayant une connaissance des données (ou plutôt des métadonnées) peuvent concevoir directement des rapports par simple glisser-déposer. Enfin, des rapports préformatés peuvent être personnalisés et adaptés en quelques clics. Concernant la relation client, les informations de Splunk sont corrélées avec les données des applications métier via la couche de services REST de la solution (API).

« Pour la visualisation des données et le reporting, nous nous contentons actuellement des outils fournis avec Splunk. Nous n'avons pas encore ressenti le besoin d'aller au-delà, car ces fonctions sont très riches,» assure Rached Tayech.

Toutefois, certains éléments d'infrastructure sont à considérer ou à tester. *«Dès le début du POC, et*

comme l'indique clairement l'éditeur, nous avons rapidement constaté de gros problèmes liés aux entrées-sorties dans les environnements virtualisés. D'où la nécessité d'installer la solution sur des serveurs physiques,» conseille le responsable Big Data.

Pas Hadoop, et pas besoin de data scientist

L'entreprise utilise Splunk, et n'a pas encore trouvé d'intérêt pour Hadoop, même si cela est possible. «D'ailleurs, nous menons déjà une réflexion pour savoir quel stockage (Data Lake) nous choisirions de favoriser dans le cas où nous déploierions Hadoop en plus de Splunk, afin de ne pas dupliquer les informations,» confie Stéphane Duarte.

Côté compétences techniques, on évoque régulièrement le besoin de data scientist dès qu'il s'agit de Big Data. «Pas besoin avec Splunk qui apporte un premier niveau de découverte automatisée des schémas de données. Ensuite, les informaticiens peuvent définir des métadonnées pour réaliser un mapping intelligent, sans aucun codage,» précise Rached Tayech. «Un informaticien connaissant les données et la BI ayant une formation de type universitaire intégrant des mathématiques et des statistiques sera totalement opérationnel après quelques jours de formation.»

Grâce à la centralisation des données de la solution et les premiers POC pour réunir les données système, deux personnes suffisent à e-Commerce Auchan pour prendre en charge l'environnement Splunk.

Le déploiement du projet complet sera réalisé dans les semaines à venir. Mais déjà Stéphane Duarte annonce: «Depuis que nous avons effectué des démonstrations de Splunk aux métiers, les demandes se multiplient alors que nous avons à peine commencé ! »

A lire aussi :

[Silicon Valley Tour – Apporter de la visibilité dans le capharnaüm applicatif](#)

[Larry Ellison : un faux départ et de vraies questions pour Oracle](#)

Crédit Photo : Dooder-Shutterstock