

Sébastien Carriot, AT Internet: « gérer, transformer et sécuriser 1 Po de données par mois »

Directeur technique d'AT Internet, un spécialiste européen du web analytics et de la data intelligence, Sébastien Carriot et son équipe d'informaticiens doivent gérer chaque mois un flot de plus de 300 milliards d'informations. Quels sont ses choix d'infrastructure informatique ? Quels ont été les développements logiciels ? Quels sont ses enjeux business des prochaines années ? Que retient-il de ses près de 20 ans de direction informatique ?

Silicon.fr : AT internet est un spécialiste européen du web analytics. Concrètement, quelles sont les problématiques IT de votre entreprise, quelles sont vos missions ?

Nous sommes effectivement dans le métier du [web analytics](#) et plus globalement de la data intelligence. Nos clients sont des grands comptes tels que LeBonCoin ou Deutsche Telekom pour qui nous traitons de très gros volumes de données de l'ordre de 300 milliards d'informations par mois. Cela représente un volume de 1 Po de données, qui s'enrichit de quelques dizaines de To tous les mois.

Notre problématique est triple. Nous devons non seulement gérer ces gros volumes de données, avec un accès temps réel tant pour l'insertion que pour la consultation des données, mais également transformer ces données brutes en « informations » – le plus souvent contextualisées-, pour créer des indicateurs de performance pertinents pour nos clients. Je dois enfin veiller à l'intégrité de ces données et de ces calculs dans une véritable démarche de « qualité ». Je n'aime pas trop l'expression mais cela ressemble au Big Data.

Comment a évolué votre infrastructure matérielle ces dernières années ?

La société a été créée en 1996 et nous avons historiquement misé sur un parc informatique homogène de serveurs 1U [Xseries](#) IBM. Mais nous avons rapidement fait preuve de pragmatisme, en choisissant au fil des années la meilleure technologie pour répondre à nos besoins dans la performance, la disponibilité ou le stockage.

Aujourd'hui, notre infrastructure s'appuie sur 350 serveurs, hébergés dans un datacenter SFR près de Bordeaux, et comprend toujours des Xseries IBM ou des petits serveurs Intel x86, que nous utilisons pour la collecte des données. Pour le processing des données et l'ETL, nous avons par contre opté pour des machines beaucoup plus puissantes équipées de cartes [Fusio IO](#) de 1 To. Enfin pour la partie bases de données, nous utilisons des serveurs FastTrack (HP/Microsoft) sous SQL server. Chaque machine a un coût de revient de près d'un million d'euros sur 3 ans mais elles correspondent pleinement à nos besoins ! Enfin nous utilisons des serveurs avec des puces octo-coeurs pour des bases de données, avec SAN fibre pour héberger les datamarts.

Outre cette infrastructure, nous utilisons également les solutions Cloud [Microsoft Azure](#) et [Amazon AWS](#), en particulier pour gérer le *front office* de nos applications, et les rapprocher de nos clients, ou

faire face à des pics de charges ponctuels et faciles à anticiper comme le reporting mensuel. Ces offres ont l'avantage d'être simples, scalables et financièrement très lisibles. Par contre, cela coûte relativement cher en bande passante et cela pose de véritables problèmes de confidentialité à certains de nos clients, en particulier depuis le Patriot Act et les différents scandales impliquant la NSA...

Avez vous également mis au point des solutions logicielles spécifiques pour gérer une telle infrastructure ?

L'essentiel de notre travail consiste à gérer l'hétérogénéité du parc informatique. Nous avons une équipe de 15 personnes qui travaillent sur la virtualisation de toute notre infrastructure de stockage, de sorte qu'elle puisse fonctionner sur n'importe quel serveur, chez nous ou dans le cloud, qu'on soit sur une base SQL ou une orientation vers Hadoop.

Historiquement, ces développements s'appuient sur des solutions Microsoft comme .Net ou SQL server, que nous trouvons efficaces et en phase avec nos besoins de « time to market ». Nous avons aussi testé Oracle et ses machines Exadata mais cet environnement nous a semblé par contre plus lourd.

Nous avons également 8 développeurs sur le front, qui travaillent aussi bien au développement de notre interface client, le célèbre Xiti, mais également sur l'intégration des données dans des progiciels de Business Intelligence ou de CRM, pour construire leurs tableaux de bord métier.

Quel sera votre sujet IT prioritaire des prochaines années ?

Le parc informatique des internautes évolue à grande vitesse avec la montée en puissance de nouveaux devices tels que les smartphones, les tablettes, les consoles, les téléviseurs connectés et au-delà ce qu'on appelle « l'internet des objets ».

Il faut non seulement traiter cette hétérogénéité du parc et des systèmes d'exploitation, mais également modifier nos modes de collecte des logs et les informations que nous partageons avec nos clients.

Nous avons par exemple intégré les données du gyroscope du smartphone ou d'une tablette, et nous sommes désormais capables d'indiquer à nos clients si leurs utilisateurs consultent leur application en mode portrait ou en mode paysage. Ce type de question ne se posait évidemment pas avec un PC...

Et ce type de problématique va également se poser avec les nouvelles interfaces machines comme les écrans tactiles, les systèmes de reconnaissance audio comme Siri ou gestuelle comme Kinect.

C'est un véritable défi technologique, en particulier dans les pays émergents qui sont rarement équipés de PC, et qui utilisent internet principalement depuis leur smartphone, avec des usages parfois beaucoup plus avancés qu'en Europe ou aux Etats-Unis...

Avec le recul, quel conseil pourriez vous donner à d'autres directeurs informatiques ?

Je leur donnerais deux conseils. Le premier c'est de ne pas être aveuglé par la technologie en voulant se faire plaisir. Les directeurs informatiques sont technophiles et croient parfois qu'une

nouvelle technologie peut résoudre tous les problèmes d'un coup de baguette magique.

Le second conseil, qui est d'ailleurs lié au premier, est ce que j'appelle la « dette technique ». Les nouvelles technologies ne remplacent jamais vraiment les anciennes et, avec les années, on se retrouve bien souvent à gérer en parallèle plusieurs générations de parcs informatiques ce qui pose des problèmes techniques, financiers mais également managériaux en allongeant les temps de formation.

Avec les années, on apprend donc à être pragmatique, à ne pas tout attendre de la technologie. L'essentiel à mes yeux c'est d'écouter le client final en lui proposant les solutions les plus réalistes.

Lire aussi :

[Mathieu Llorens \(AT internet\) : « la donnée web est au cœur du système d'information de l'entreprise »](#)