

Big Data : Blablacar copilote sa BI avec HP, Tableau et Dataiku

Blablacar est connu dans la sphère des start-ups pour avoir récemment levé 100 millions de dollars. Le concept de la société est simple : mettre en relation des personnes pour un service de covoiturage. Et le succès est au rendez-vous avec plus de 10 millions de membres dans 13 pays (avec l'ouverture récente du Brésil et de l'Inde). « *Au total, nous avons 1 millions de membres supplémentaires toutes les 6 à 8 semaines* », explique **Gaëlle Périat**, Data Nerd chez Blablacar. Et ces membres discutent entre eux, car la jeune pousse a tissé, en plus de son service, une communauté sur les réseaux sociaux et joue à fond cette carte y compris pour ses campagnes marketing.

Face à cette croissance rapide, l'équipe dirigeante a très vite compris l'importance de l'analyse de données pour améliorer l'expérience client et a donc chargé **Nicolas Menoux**, consultant IT et ancien DSI, de la responsabilité d'une équipe en charge du projet BI. Arrivé pendant l'été 2014, il a accompagné l'adoption d'un outil de BI taillé pour le développement exponentiel des informations (trajets, lieux, kilométrages, tarifs, notations, profils, ancienneté, etc) . « *L'objectif était d'intégrer de la BI sur les données de production reposant sur une base de données MySQL. Auparavant, l'extraction de données comportait des risques de pertes d'informations avec les mises à jour sur des tables devenues trop grosses* », constate Nicolas Menoux. Le choix s'est porté sur **Vertica de HP** sur un **cluster Hadoop Cloudera**.

Techniquement, le cluster repose sur 4 machines tournant sur Debian 6 avec 64 Go de RAM et 4 To de stockage. La solution Vertica a d'abord été installée sur 2 VM sous Debian. « *Nous allons basculer sur l'édition Entreprise de Vertica avec une capacité de 3 To de stockage* », assure Gaëlle Périat. Une capacité jugée suffisante pour encaisser les estimations de croissance annuelle de la société. La première étape après l'intégration de Vertica a été de l'alimenter avec la récupération des données des abonnés depuis 2005. Une opération qui s'est déroulée sans difficultés, selon la team Big Data de Blablacar.

Dataiku permet « de partir en vacances »

Pour autant, tout ne repose pas uniquement sur la seule offre de HP, la start-up s'est adjointe les compétences de **Tableau Software** pour la partie data visualisation et reporting. L'objectif était de donner aux métiers des éléments visuels compréhensibles et collant à leurs besoins. Un but similaire pour les requêtes : « *avant nous avions des requêtes Excel/Excel, aujourd'hui les requêtes sont automatisées avec des rapports quotidiens pour les équipes marketing. Nous pouvons aussi faire des requêtes plus fines et plus granulaires pour les métiers* », indique Gaëlle Périat. Cette automatisation des tâches a été confiée à l'offre Data Science de **Dataiku**, une plateforme d'analyse de données et de construction d'applications prédictives. « *Nous avons réalisé un POC (un prototype, NDLR) en mai dernier avec [Dataiku](#) et nous l'avons implanté rapidement* », poursuit la Data Nerd. Elle a été séduite par la plateforme où « *des recherches peuvent être créées à partir d'un simple script Python* ». Un brin provocatrice, elle admet que cet outil lui permet « *de partir en vacances* », preuve que l'automatisation n'est pas un vain mot.

Les bénéfices de cette stratégie sont donc au rendez-vous avec une plus grande rapidité et fluidité dans les requêtes et les rapports pour les métiers. Les campagnes marketing sont plus ciblées et plus réactives, notamment sur les réseaux sociaux comme Facebook où Blablacar est très présent. Elles prennent surtout en considération la montée en puissance de la start-up avec des marchés aussi importants que l'Inde ou la Russie. Les jeux de données vont s'enrichir au fur et à mesure du développement des services. « *Nous avons une solution hyper scalable* », juge Gaëlle Périat. Les relations avec les métiers ont changé, « *il n'y a plus de prises de décisions sans les data* », assure-t-elle.

Des ambitions dans le temps réel et le machine learning

L'équipe Big Data fourmille d'idées pour faire évoluer son architecture. En premier lieu, Nicolas Menoux a une idée fixe : poursuivre la « *'nonification' des bases de données SQL existantes vers Cassandra* » (autrement dit, le passage graduel aux technologies NoSQL). L'équipe mène en parallèle des tests [de Spark, le concurrent de MapReduce](#) sur une partie de Cassandra pour obtenir des fonctions analytiques qui **se rapprochent du temps réel**. « *L'idée est de pouvoir en faire une zone de test pour de nouvelles choses* », confie Gaëlle Périat.

Autre point d'amélioration attendue, un renforcement du **cluster Hadoop sur 3 nœuds**, la possibilité de réaliser **des backups dans Hadoop** et le test de l'offre Big Data Haven de HP. Sur l'automatisation des tâches, « *nous allons monter sur la partie machine learning avec Dataiku en 2015* », assure la Data Nerd. L'équipe Big Data de Blablacar n'a donc pas de quoi chômer et continue à recruter. Avis à ceux qui seraient tentés par y faire un bout de chemin...

A lire aussi :

[Big Data : diagnostiquer les troubles psychologiques en scrutant Twitter](#)

[Le Big Data toujours aussi bouillonnant dans la Silicon Valley](#)