

Romain Chaumais, Ysance : « le Big Data en temps réel n'est pas une exigence, c'est une libération »

Silicon.fr : Les technologies Big Data s'orientent de plus en plus vers le temps réel. L'offre est-elle prête ?

Romain Chaumais (Ysance) : Oui, je le pense. Dans le monde Hadoop, les évolutions sont très rapides. Début 2015 encore, on ne pouvait pas se permettre de se focaliser uniquement sur Spark (un framework Open Source de calcul distribué, fonctionnant en mémoire vive, NDLR). Un an après, tous nos nouveaux projets s'appuient sur cette technologie. Dans un an, on verra sortir de vrais cas d'usage avec des technologies temps réel. En fait, le nombre de scénarios métier ayant réellement besoin du temps réel n'est pas si élevé, mais développer à 100 % en Spark permet de mettre en place un seul flux de calculs tant pour le batch que pour les applications temps réel. C'est une simplification des architectures, au prix, certes, d'un léger surinvestissement en infrastructure, Spark étant gourmand en mémoire.

Avez-vous des exemples des bénéfiques qui peuvent en découler ?

Pour un gros industriel travaillant dans le domaine de l'agriculture, Ysance a récemment travaillé à la réécriture d'une application permettant de calculer les quantités d'engrais à apporter à une parcelle en fonction de la météorologie. Un calcul critique car la molécule employée est météo-sensible. Sur une grosse infrastructure SQL Server Fast Track, le calcul demandait une semaine ! Avec une application développée en langage Scala sur Spark, et hébergée chez AWS (le Cloud d'Amazon, NDLR), le même calcul ne demande plus que 1 minute 26 secondes. Pour quelques centaines d'euros de coûts d'infrastructure ! Cette réduction drastique change tout bonnement le modèle économique de cet industriel.

Je pense aussi à une entreprise du secteur de l'automobile fabriquant des injecteurs diesel. Chacune de ces pièces vaut 150 euros, dont environ 100 euros de coût de matière. L'usine qui les produit réalise quelque 1 500 tests de qualité et entre 5 à 7 % de la production est déclarée non conforme. Sur ce type de configuration, parvenir à corrélérer en temps réel les événements sur la chaîne de production avec les rejets représente un gain énorme par rapport au mode batch. On peut arrêter la chaîne immédiatement et on évite les rappels de produits déjà livrés aux clients.

Avec le couple Spark – Scala, on découvre que l'approche ensembliste qui a prévalu pendant des décennies peut être dépassée par une nouvelle approche fonctionnelle. C'est un peu la 3^{ème} génération d'applications Big Data, après celle des clusters Hadoop – MapReduce, où il fallait écrire en Java, puis celle marquée par l'arrivée de Hive, qui a permis de réintégrer les compétences SQL. Cette 3^{ème} génération ne fera pas de concession sur les compétences. C'est pourquoi les profils Spark-Scala sont si recherchés : si je pouvais en embaucher plusieurs aujourd'hui même, je le ferais sans hésiter.

Cette réécriture concerne-t-elle seulement les nouvelles applications ou va-t-elle toucher aussi l'existant ?

R.C. : Je ne vois aucune raison, aujourd'hui, de modérer ma réponse. Prenez l'exemple de la coopérative agricole que je citais plus haut, où le prototype a été monté en à peine un mois malgré la traduction des règles métier du SQL vers Scala. Par ailleurs, je ne vois plus une start-up qui ne code pas sur Spark. Le DSI de Chronopost, qui avait choisi cette technologie, vient aussi d'être récemment choisi pour mener la même transformation au sein de la branche courrier de La Poste. Le temps réel n'est pas une exigence, c'est une libération. Cette transition permet d'arrêter de se poser sans cesse des questions techniques.

Ysance a été fondé en 2005 par deux ingénieurs, Laurent Letourmy et Romain Chaumais. La société de services devrait dépasser les 11 millions d'euros de chiffre d'affaires en 2016. En septembre dernier, Ysance a annoncé avoir réalisé une levée de fonds de 5 millions d'euros auprès de Creadev, la société d'investissement entrepreneuriale de la famille Mulliez. Un apport en capital qui vise à développer une offre d'édition de logiciels, autour du Data Management Platform (DMP).

A lire aussi :

[Big Data : Mappy accélère son cluster Hadoop... sans acheter de serveurs](#)

[Mike Olson, Cloudera : « dans Hadoop, temps réel et batch sur les mêmes données »](#)