

E. Brochard, Pentaho : « croiser les données à la source, en temps réel »

L'éditeur de solutions analytiques (y compris mobile et Big Data) Pentaho fête ses dix ans, et vient tout juste de débaucher le patron de la stratégie Business Analytics d'IBM au poste de directeur produit. Rappelons que Pentaho s'est bâti sur le modèle de l'open source professionnel et propose une version libre (plate-forme complète de BI – Community Edition) et une version payante (Enterprise Edition). Interview avec Emmanuel Brochard, le directeur général pour la France, croisé sur le salon Big Data 2014, qui s'est tenu la semaine dernière à Paris.

Silicon.fr : Pentaho à 10 ans. L'occasion de faire le point sur votre entreprise et ses leviers économiques...

Emmanuel Brochard : Nous assemblons des briques de communautés open source, dont nous avons le plus souvent recruté les animateurs. Pentaho compte 350 employés dans le monde, et s'est installé en France en 2010. La filiale compte 6 collaborateurs, et nous recrutons en 2014 des profils commerciaux, marketing et avant-vente.

Les zones EMEA (Europe, Moyen-Orient, Afrique) et APAC (Asie Pacifique) emploient 25% des employés de la société et réalisent 35 % du chiffre d'affaires. Nos revenus, en croissance de 80% par an, sont tirés par la BI traditionnelle (50% de notre activité), qui croît de 25%, l'analytique embarquée (25% du CA), en hausse de 25%, et le Big Data (déjà 50% du CA), qui enregistre une croissance de plus de 100%. Depuis 2013, nous constatons un inversement de la tendance. En effet, la BI traditionnelle se maintient, mais on assiste un remplacement par des technologies plus innovantes et moins chères.

Comment se différencier avec une offre open source sans emprisonner le client ?

Pentaho propose une couverture décisionnelle complète avec l'intégration et l'analyse de données, l'accès aux informations (ETL), le reporting, le datamining, la visualisation, le Big Data, l'analyse prédictive... Pour offrir un maximum d'ouverture et éviter les bases propriétaires, nous ne proposons pas de référentiel spécifique, mais un catalogue de métadonnées qui se connecte à des référentiels.

De même Pentaho dispose d'un maximum de connecteurs natifs vers les SGBDR, mais aussi Hadoop, NoSQL (MongoDB), etc. Un effort qui permet d'offrir un maximum d'ouverture en affichant les meilleures performances. Enfin, l'offre est modulaire, et chaque module peut s'intégrer avec l'existant de l'entreprise sans avoir à tout installer. Cependant, la cohérence de la plate-forme Pentaho offre le meilleur résultat.

Vous surfez actuellement sur la vague du data blending (croisement des données). En quoi Pentaho est-il différent de Tableau Software ou autres ?

La version 5 de Pentaho a été totalement ré-architecturée pour intégrer le Big Data. La plate-forme n'a pas été adaptée, mais remaniée nativement. Ce qui a également permis d'intégrer le data blending à la source. Les solutions du marché mettent généralement en corrélation diverses

sources de données à travers le logiciel client, dans le tableur Excel par exemple.

Pentaho croise les données à la source : SGBD, Big Data, réseaux sociaux, réseaux de télécommunication... Notre ETL PDI (Pentaho Data Integration – Kettle à l'origine) réalise le croisement ou mélange de données à la source, sans stockage et en temps réel. Et cela aussi bien depuis un datawarehouse que depuis les données de production, par exemple. Et il est possible d'intégrer des mécanismes de qualité dans le schéma de données. Par exemple, l'entreprise peut appliquer des règles de priorité sur la bande passante à des requêtes à certaines heures, selon divers critères.

Le data blending permet de détecter des opportunités inattendues, ou de proposer des offres contextuelles très pertinentes. Ainsi, croiser les données d'un distributeur de billets et du datawarehouse lorsque le client retire de l'argent permet de lui proposer à un moment opportun des produits financiers avec un profil bien qualifié.

Cette complexité reste-t-elle à la portée du programmeur ou de l'administrateur de données ?

Pentaho est la plate-forme la plus lisible et la plus 'user friendly' du marché. Un maximum d'opérations s'effectue de manière totalement graphique et sans codage. Y compris dans l'ETL qui ne génère pas du script, mais des paramétrages. Et il en va de même pour les outils de qualité des données.

Qu'en est-il de votre réseau de partenaires sur le terrain ?

Actuellement, nous adressons en direct les entreprises du mid-market, mais recrutons aussi des cabinets de conseils ou grands intégrateurs comme Capgemini, Sopra, Business et Décision, Micropole... Par ailleurs, nous avons signé en 2012 un accord pour intégrer Pentaho Business Analytics for Big Data dans les solutions Cisco UCS. Et nous constatons actuellement une demande croissante sur ce segment "embarqué" (OEM).

En complément

[François Bourdoncle, Exalead : « la guerre commerciale du Big Data a déjà commencé »](#)

[P. Livaudais, ParStream : « conjuguer les volumes de Hadoop et la vitesse du In-Memory »](#)