

[Xavier Guérin, MapR : « Les projets Big Data évoluent vers la maintenance prédictive ou les objets connectés »](#)

Xavier Guérin, vice-président Europe du Sud et Benelux, revient pour Silicon.fr sur les différentes actualités de MapR, présence en France, développement des partenariats, mais aussi l'implication financière de Google dans le spécialiste du Big Data.

Silicon.fr : Combien d'employés compte MapR France un an et demi après l'ouverture du bureau parisien ?

Xavier Guérin : Pour le moment, nous avons surtout investi sur MapR Benelux en recrutant une équipe commerciale et avant-vente. En France, nous lançons des recrutements supplémentaires pour début 2015 : des commerciaux et des data engineers/data scientists (des personnes capables d'évaluer des datasets et de déterminer quel type d'outil peut apporter de la valeur ajoutée au client). Au niveau mondial, les choses avancent vite, puisque de 240 employés début 2014, nous serons plus de 350 avant la fin de l'année, avec un fort investissement en EMEA Sud.

Où en est le réseau de partenaires dans l'Hexagone ?

Nous disposons d'un réseau de partenaires certifié et mature, de deux types.

Des intégrateurs Hadoop peuvent accompagner une entreprise dans la mise en œuvre d'une application ou d'une infrastructure Big Data : Ysance, Octo, Fast Connect (Bull), Celeris, Affini-tech et Hurence.

Par ailleurs, des sociétés de conseil peuvent intervenir en amont après de dirigeants : Blue Stone, Umanis, CGI/Logica et Edis.

Un point sur les clients français ? Et sur les projets réellement déployés en production ?

Nos clients (comme ceux de nos concurrents) restent frileux sur la communication des projets Big Data. Client de MapR, Cdiscount s'est publiquement félicité d'avoir intégré MapR sur son site de production à Bordeaux, avec un site de reprise sur incident (Disaster Recovery) sur Paris. En effet, MapR est le seul à proposer une solution Hadoop incluant la fonction de Disaster Recovery.

Parmi nos autres clients, une banque et un grand de la distribution ont validé MapR et sont sortis du POC (proof of concept ou prototypage évolué). Ils seront en production d'ici à la fin de l'année.

Depuis la dernière manifestation Big Data Paris (début avril 2014), une vingtaine de POC sont en cours. Évolution notable : ces projets sortent des verticaux « classiques » (distribution, Finance, télécommunications...) pour évoluer vers de la maintenance prédictive ou des objets connectés.

MapR a levé [80 millions dollars en juillet dernier](#), qu'allez vous en faire ?

En fait, la levée de fonds n'était pas une priorité absolue pour MapR, car nous disposons d'un cash-flow équilibré. Cependant, le marché a beaucoup bougé, et nos concurrents ont levé des millions pour accélérer leur développement et leur expansion. Nous avons donc monté ce tour de table afin

de favoriser l'accélération du développement des produits (dans nos centres de R&D dans la Silicon Valley et en Inde), notre expansion et le renforcement des filiales en place.

Pourquoi Google a-t-elle choisi d'investir sur le pionnier d'Hadoop dont le modèle économique est le moins Open Source ? Après avoir aussi investi aussi dans Cloudera.

Il faut bien distinguer les deux investissements et les entités de Google ayant investi.

L'entité Google Ventures (placements financiers) a effectivement investi dans Cloudera. Il s'agit bien de placements plutôt spéculatifs. Tandis que Google Capital qui sélectionne un nombre très limité de sociétés (moins de 10 aujourd'hui) a investi dans MapR. Cette entité de Google mise sur des partenariats et les investissements à long terme, avec un représentant de Google Capital au comité d'administration de MapR. À noter : nos investisseurs historiques ont tous réinvesti environ 30 millions d'euros, confirmant leur confiance.

D'où proviennent vos revenus et selon leur répartition ? Que contiennent vos licences ?

Nos revenus sont répartis de la façon suivante : 90% du chiffre d'affaires provient de la vente de licences et 10% de la formation et des services. MapR propose deux types de licences. La version communautaire gratuite regroupe toutes les fonctions (y compris l'administration) sauf la haute disponibilité.

La version entreprise est proposée sous forme d'achat de licence annuelle ou trisannuelle incluant usage, mises à jour et support 24x7 illimités, tout inclus sans options. (2 900 euros par nœud).

Nos services sont la formation/certification, et des prestations de consulting à travers les partenaires avec des data engineers ou data scientists très spécifiques. Nous intervenons alors en appui, en accompagnement des partenaires.

Côté formation, la quasi-totalité des contenus et des évaluations est en ligne : supports, QCM, vidéos... et certifications.

Nous avons déjà un partenaire formation au Royaume-Uni, et nous menons actuellement des discussions ne ce sens avec certains spécialistes français.

Hortonworks a opté pour le 100% Open Source, Cloudera également. Pourquoi MapR maintient-elle des briques propriétaires ?

Outre la haute disponibilité dont nous venons de parler, comment réalise-t-on aujourd'hui une application temps réel Big Data, par exemple ? Car le système de gestion de fichiers HDFS (socle de Hadoop) n'est pas conçu pour le temps réel. Ainsi, il est impossible d'obtenir un snapshot consistant, ou de déployer une reprise sur incident (Disaster Recovery) digne de ce nom. MapR FS le permet et sait même faire cohabiter Map Reduce 1 (MR1) et Yarn (MR2) sur un même cluster.

Dans l'architecture Yarn, HDFS et son single namenode subsistent. Et même s'il est sécurisé avec un second single namenode, la haute disponibilité s'avère très relative.

MapR revendique cependant un positionnement Open Source volontariste...

Absolument. Par exemple, nous avons initié et nous contribuons au projet Apache Drill avec une dizaine d'employés MapR dédiés. Cette brique vise à apporter le temps réel avec accès multiples à diverses sources de données. Ainsi, il devient possible de lancer une requête jointe en SQL ANSI, avec ou sans modèle relationnel. Y compris sur un fichier texte, et quelle que soit la taille des fichiers. (Pour les développeurs: avec schémas classiques HCatalog dans Hive ou -sans s'en

préoccuper- directement dans le fichier). Drill offre donc la possibilité de requêtes interactives, sans oublier la capacité de structurer le système de gestion de fichiers pour restreindre et contrôler les accès à tous les contenus.

A lire aussi :

[Big Data : YARN remplace MapReduce et apporte la simultanéité à Hadoop 2.0](#)