

MapR : et si le Big Data entrerait en bourse ?

MapR envahit rapidement l'Europe...

John Schroeder – En 2009, j'ai cofondé MapR avec MC Srivas (aujourd'hui directeur technique), impliqué dans les équipes de Google qui utilisaient les technologies comme MapReduce, Big Table ou GFS depuis 1998.

Nous avons lancé notre première solution en 2011, et aujourd'hui nous sommes présents aux États-Unis, au Royaume-Uni, en France, en Allemagne, en Suède, au Japon, à Singapour... Et nous poursuivons notre expansion.

Nous avons déjà recruté des talents confirmés qui viennent de sociétés comme Microsoft, EMC, Google, etc.

D'ailleurs, Michael Lehman *[NDLR : ex-directeur financier de Sun Microsystems, racheté par Oracle en avril 2009]* vient tout juste de rejoindre notre conseil d'administration, afin de nous apporter son expérience et nous préparer à une introduction en bourse. Cela nous permettra d'accélérer notre développement et de conforter notre expansion internationale.

Silicon.fr : comment positionnez-vous votre solution Big Data ?

Notre plate-forme Hadoop comporte des outils évolués d'administration et de supervision permettant de simplifier le déploiement et la maintenance de la solution dans les entreprises.

En outre, nous avons ajouté notre propre technologie de stockage (déposée) qui enrichit l'ensemble de fonctions de haute disponibilité et de reprise après incident, entre autres.

Utilisateur des briques Open Source Hadoop (du projet Apache), MapR est également un grand contributeur en reversant beaucoup du code que nous développons à la communauté.

Outre Hadoop et notre plateforme, nous avons également une base de données NoSQL *[NDLR : not only SQL]* conçue pour et intégrée à Hadoop pour favoriser le développement d'applications critiques. Et notre plate-forme M7 apporte la haute disponibilité aux applications Hadoop et Hbase.

De plus, MapR a initié le projet Apache Drill qui ouvre Hadoop au SQL ANSI *[NDLR : et donc aux SGBD traditionnels]* et au travail avec les données en temps réel. En effet, HCatalog répond à de nombreux besoins, mais pas à l'accès aux données en temps réel, entre autres.

Comment procédez-vous pour trouver des

spécialistes Hadoop dans les différents pays ?

Il est vrai que le recrutement de spécialistes n'est pas simple, mais de nombreux informaticiens sont attirés et séduits par ces nouvelles technologies. Ce qui simplifie grandement les choses. Rappelez-vous la situation d'Oracle au début des années 80...

En outre, nous proposons des formations et des certifications à des intégrateurs et des spécialistes de plus en plus nombreux. Et nous sélectionnons comme partenaires des sociétés très pointues sur ces sujets.

La rupture technologique Big Data impacte-t-elle aussi les développeurs ?

Pour les développeurs, cela ne change pas grand-chose, mais ils profitent des avantages et de la performance d'une infrastructure massivement parallèle.

Ils rédigent leur code, et la plate-forme s'occupe du reste : Map distribue les traitements et les données sur les divers nœuds, et Reduce assemble les résultats pour apporter la réponse. Enfin, Hive apporte un langage de type SQL [NDLR : HiveQL] pour manipuler les données stockées sur une plate-forme Hadoop.

Quels sont vos atouts face à vos deux concurrents, également pionniers d'Hadoop : Cloudera et Hortonworks ?

Il est vrai que nous utilisons tous les briques Hadoop Open Source du projet Apache. Et tout le monde a intérêt à préserver l'interopérabilité entre les distributions Hadoop, afin de ne pas reproduire les problèmes d'interopérabilité et de migration que l'on rencontre entre Oracle, IBM DB2, Sybase, Microsoft SQL, etc.



Aujourd'hui, les entreprises ne veulent plus être captives des technologies d'un éditeur.

Par ailleurs, nous proposons tous du support et des formations sur nos distributions.

Contrairement à Hortonworks, nous proposons des outils d'administration et de supervision sur notre plate-forme.

En outre, nous proposons une technologie de stockage qui simplifie grandement les traitements et l'interopérabilité d'Hadoop avec l'existant des systèmes d'information, un peu comme peut le faire Teradata [NDLR : et pas Cloudera...].

Et bien entendu, tout cela est très documenté, les fonctions, les API, etc.

En termes de connecteurs et d'intégration avec les autres, l'interopérabilité facilite grandement les choses, puisque nous utilisons les mêmes briques de base.

Quelles autres tendances technologiques voyez-vous le plus associées au Big Data ? Le cloud ? Public ou privé ? ...

Les entreprises s'intéressent un peu au cloud public pour Big Data, mais globalement elles optent surtout pour une utilisation en cloud privé.

En effet, elles préfèrent maîtriser les infrastructures et pouvoir les optimiser, tout en conservant les informations dans leur propre système d'information. Aussi bien pour la confidentialité que pour les soucis de proximité et de performance.

On constate que la combinaison virtualisation-Hadoop est une tendance qui s'accroît actuellement, et qui redouble l'intérêt de nombreuses entreprises.

Assez logique en somme.

Voir aussi

[Quiz Silicon.fr - Le vocabulaire du cloud](#)