

# Big Data : La Silicon Valley a toujours le béguin pour Hadoop

Au cours de la tournée auprès de “startups applicatives” dans la Silicon Valley, nous avons une fois encore constaté combien la plupart d’entre elles optent pour des solutions open source, dont une grande partie pour Hadoop. Et le succès des trois pionniers (Cloudera, Hortonworks et MapR) a clairement accéléré le phénomène.

## **Atscale ouvre Hadoop à la BI traditionnelle**

Comment rendre les données Hadoop accessibles à tout utilisateur via un simple navigateur Excel ou son outil de Business intelligence traditionnelle (Cognos, Microstrategy, Qlik, Tableau...)?

Le credo d’Atscale : proposer de la “BI sur Hadoop”. *«La solution attendue doit répondre à deux attentes fortes et légitimes : Les utilisateurs métiers souhaitent plus de liberté et du self-service pour accéder aux applications analytiques, et les informaticiens attendent un contrôle efficace des données et une cohérence globale (en ligne avec le besoin de fiabilité exigé par les métiers),»* explique David Mariani, cofondateur et CEO d’Atscale.

C’est pourquoi la société a conçu son Intelligent Platform qui s’installe au-dessus d’une plateforme Hadoop, afin de rendre ses informations accessibles très simplement aux outils de Business Intelligence traditionnels. *«La plateforme intercepte les requêtes SQL (ou MDX pour Excel) et les interprète pour interroger HDFS dans le langage le plus approprié. Puis il enrichit les données selon un schéma XML prédéfini. Les cubes Olap générés sont réalimentés au fur et à mesure,»* précise le CEO. *«Toutefois, notre serveur ne stocke que les définitions et les métadonnées, et pas les données elles-mêmes.»*

Dans sa version 3.0 disponible depuis fin octobre, Atscale Intelligent Platform intègre la technologie auto-apprenante Adaptive Cache. Cette dernière analyse les modèles de requête pour mettre en place un cache intelligent diminuant encore les temps de réponse.

Créée en 2013 par David P Mariani (CEO, ex de Yahoo et de Klout) et Matthew Baird (CTO, ex d’Oracle et Siebel), Atscale a levé 2 millions de dollars en 2013 et 7 autres en juin 2015.

## **Solix propose l’archivage des données sous Hadoop**

Avec 175 employés, plus de 6000 clients et plusieurs bureaux dans le monde (USA, Royaume-Uni, Australie, Inde et Dubai), Solix n’est plus vraiment une start-up. En revanche, elle innove en utilisant Hadoop de manière plutôt innovante : pour l’archivage long terme des informations d’entreprise.

Ce spécialiste de la gestion du cycle de vie des informations d’entreprise (ILM pour Information Lifecycle Management),

*« L’archivage consiste à migrer, protéger et indexer des données sur un équipement secondaire afin de réduire les coûts, optimiser l’infrastructure informatique, proposer sécurité et gouvernance des données, faciliter l’accès à l’information et favoriser le respect des obligations réglementaire. A ne pas confondre avec la sauvegarde ou l’ETL,»* analyse Vikram Gaitonde, vice-président Produits chez Solix Technologies. *«Apache Hadoop est en open source et offre une infrastructure massivement évolutive, avec des traitements en parallèle. De hautes performances sur des équipements banalisés et à bas coût gérant les données structurées*

*ou non. Un téraoctet de données sous Hadoop coûte plus 55,5 fois moins cher que sur un environnement en production.»*

Hadoop a donc été retenue pour une partie conséquente de l'archivage dans le cycle de vie des données: "les archives froides". L'éditeur propose Solix Enterprise Management Suite pour archiver les applications "à la retraite" et leurs données, améliorer les performances des SGBD (en les déchargeant dynamiquement des données les moins accédées), prendre en charge des données de test, etc. En complément, Solix conjugue ILM et Hadoop dans sa Big Data Suite pour archiver les données structurées et non structurées peu ou plus utilisées. Sa forte connaissance des applications et des SGBD permet de proposer des solutions très intégrées et optimisées. La solution est déjà adoptée par un spécialiste de l'aérospatiale/Défense qui a constaté que 25% de ses applications et plus de 80% des informations non structurées étaient inutilisées. *«Cette suite sous Hadoop a permis d'optimiser l'infrastructure existante à un coût record,»* conclut Vikram Gaitonde.

### **Hortonworks rentable dès 2017**

Depuis son [entrée en bourse réussie en décembre 2014](#), Hortonworks accélère. Cela s'est traduit par le [rachat de Sequence IQ](#) et ses automatisations de déploiement de clusters Hadoop en avril, suivie du [rachat d'Onarya](#) pour l'Internet of Anything (IoAT) en août, et par l'installation du siège de l'entreprise dans tout un immeuble de Santa Clara (Californie).

*« Fin juin, Hortonworks comptait 556 clients dans le monde, dont 119 nous ont choisis au cours du second trimestre. Et nous comptons déjà plus de 150 partenaires,»* rapporte Herb Cunitz, président d'Hortonworks. *«Nos revenus proviennent essentiellement du support, mais aussi de codéveloppements avec nos partenaires (comme Microsoft) ou avec nos clients. Et si notre croissance passe aussi par des acquisitions dont nous versons les technologies à la communauté open source, nous estimons atteindre la rentabilité courant 2017. D'autant plus que notre croissance s'accélère et que les clients existants recourent de plus en plus à nos services.»*

Puis le dirigeant a détaillé les quatre étapes à travers le modèle de cycle de vie Hadoop d'Hortonworks :

- **Conception** (Architect) : cette étape de conception consiste à infuser les demandes des entreprises dans le planning des évolutions itératives d'Hadoop. «Un travail qui repose sur des codéveloppements avec des partenaires (Microsoft, EMC, HP, Teradata, SAP, RedHat...), et des clients comme Target, Schlumberger, Merk ou Aetna,» souligne Herb Cunitz.
- **Développement** : il s'agit de travailler avec les communautés open source dans le cadre de l'Apache Software Foundation. Hortonworks contribue ainsi à 25 projets Hadoop, par exemple.
- **Distribution** : après avoir testé chaque version d'Hadoop il convient de la distribuer aux clients des packages avec les meilleurs composants, dans la ligne d'ODP (Open Data Platform voir [notre article](#)).
- **Support** : fournir à l'utilisateur une assistance pour déployer des clusters Hadoop, tout en identifiant les nouveaux besoins. Indispensable puisqu'aujourd'hui encore les compétences sont peu répandues.

**A lire aussi :**

[Big Data : Amazon renforce son offre Hadoop / Spark](#)

[Comment Criteo transforme Hadoop en moteur de sa rentabilité](#)

**Crédit Photo : agsandrew-Shutterstock**