

Hadoop Summit 2015 : le socle commun

ODP entre en piste

Avec plus de 1000 participants et 65 sessions dispensées sur deux jours à Bruxelles, le Hadoop Summit 2015 d'Hortonworks devient un événement européen majeur sur le Big Data.

« Déjà cotée au Nasdaq, Hortonworks n'a été créée qu'en 2011, suite à un essaimage de l'équipe de 14 ingénieurs à l'origine du projet Apache Hadoop chez Yahoo. La société compte aujourd'hui 650 employés dans 17 pays, » rappelle Herb Cunitz, président d'Hortonworks. « Nous avons séduit 332 clients dans le monde en 10 trimestres, dont 99 pendant le seul dernier trimestre 2014. »

Si les pionniers de Hadoop profitent effectivement de l'engouement des entreprises pour ces technologies, ils maintiennent leurs investissements rendus possibles grâce aux multiples levées de fonds et leur entrée en bourse.

« Parce qu'Hortonworks préserve son modèle 100% Open Source avec souscription payante à un support annuel, nous recherchons à recruter un maximum de contributeurs majeurs à Hadoop dans les divers pays où nous sommes présents. En effet, il est sain que ces spécialistes soient proches des commerciaux et des clients, et donc du terrain. Aujourd'hui 28 des 87 committers Apache Hadoop travaillent chez Hortonworks, » ajoute le président.

L'Europe représente aujourd'hui 15 à 20% à la fois de clients et des employés de l'éditeur, qui affirme poursuivre ses efforts pour faire augmenter ce pourcentage sur les deux tableaux.

ODP : juste au départ, mais bien lancé.

Une des initiatives importantes mises en vedette lors de l'événement fut ODP ou l'Open Data Platform. Comme [nous l'évoquions en juillet dernier](#), Pivotal (division Big Data d'EMC) et Hortonworks avaient décidé de collaborer sur le projet open Apache Ambari, afin de proposer une solution évoluée de provisioning, de gestion et de supervision des clusters Hadoop. En février 2015, le partenariat prenait de l'ampleur avec [la naissance du projet ODP](#) avec le ralliement d'IBM, General Electric, Hortonworks, Infosys, Pivotal et SAS, rejoints dans la foulée par Altiscale, Capgemini, CenturyLink, EMC, Splunk, Verizon, Teradata et VMware.

« ODP est aujourd'hui constitué de Hadoop 2.6 (plus précisément les projets Open Source Apache HDFS, Yarn et Map Reduce) et du projet Ambari, » précise Rob Bearden, CEO d'Hortonworks. « Notre ambition consiste à proposer aux entreprises un cœur de plateforme commun [core platform] sur laquelle peuvent fleurir les projets Big data, avec une compatibilité entre eux. Entre autres... »

Effectivement, la compatibilité entre distributions Hadoop pourrait vite représenter un frein à l'essor du Big data, voire enfermer les entreprises, comme le souligne Anjul Bhambhri, vice-président d'IBM Analytics Platform: *« Plutôt que chacun travaille dans son coin à élaborer sa propre distribution, mieux vaut unir nos efforts et collaborer à un socle commun Open Source et dans le cadre de la fondation Apache. Sur ce socle, chacun peut alors proposer ses solutions et développer des applications. D'autant plus qu'il n'est pas toujours évident pour un éditeur ou un intégrateur de suivre et d'intégrer systématiquement les nombreuses évolutions des composants Hadoop. Sans parler des fournisseurs de services et hébergeurs cloud, pour lesquels suivre les évolutions de 8 ou 10 distributions Hadoop devient vite*

compliqué. Justement, ODP permet de disposer d'un socle commun de références, favorisant l'interopérabilité des diverses solutions et applications, évitant aux entreprises de dépendre d'un unique éditeur et de ses propres choix technologiques. En outre, disposer d'une interface d'administration unique pour toutes les distributions simplifie les projets et en diminue les coûts.»

Partie prenante du projet dès son origine, IBM fait reposer son Open Platform 4.0 sur ODP, et s'engage à assurer le plus possible la compatibilité avec ses produits Big Data, de Cognos à BigInsights en passant par SPSS.

Open source, mais pas feu de camp

Pivotal, impliqué avant même le lancement de l'initiative ODP, rappelle quelques fondamentaux par l'intermédiaire de Leo Spiegel, vice-président en charge du développement et de la stratégie chez Pivotal: *«Il s'agit d'accélérer la mise à disposition d'applications Big data en proposant une plateforme centrale prévisible et avec une version homogène et cohérente pour les entreprises. Tous les développements et les évolutions d'ODP seront réalisés dans le cadre de l'Apache Software Foundation, en concordance avec les projets Open Source liés. Premier ou dernier arrivé, tout membre d'ODP dispose d'une voix lors des votes de décisions. Enfin, l'implication dans le développement, les planning et le déroulement d'ODP sont transparents et ouverts à tous.»*

Bien entendu, quelques actions sont déjà en cours, comme la constitution de quatre groupes de travail : Gouvernance and Structure, Technical Development Community, Marketing and Awareness et Apache Assist Program. Les membres cherchent actuellement à recruter un directeur exécutif et des employés pour ODP. Enfin, des discussions devraient bientôt commencer sur deux sujets: un ODP Developer Program sponsorisé (donc avec constitution d'un fonds financier) et des programmes de formation/certification pour les Universités.

La souscription pour devenir membre d'ODP est proposée à partir de 1000 dollars.

Ces quelques éléments organisationnels illustrent clairement la volonté des membres de faire d'ODP un élément stratégique de la diffusion d'Hadoop.

Toutefois, il semble raisonnable de penser que sans la participation d'acteurs comme Cloudera et MapR, le projet pourrait se heurter au scepticisme du marché et/ou des clients. Or, ces deux éditeurs mettent systématiquement en avant leurs solutions d'administration et de supervision. L'intégration d'Ambari dans ODP risque fort de ne pas les enchanter... Cependant, on notera aussi que de nombreux géants informatiques membres d'ODP sont aussi de grands partenaires de ces deux éditeurs.

A lire aussi :

[Big Data : comment SoLocal apprivoise Hadoop](#)

[Big Data : IBM, GE, SAS, Hortonworks... forment une alliance autour de Hadoop](#)