

EMC World : Greenplum ou l'ambition du data science dans le Big Data

En direct d'EMC World, Las Vegas. Au delà de l'annonce du lancement de **Greenplum Analytics Workbench**, le projet de laboratoire sur un projet de Big Data composé de 1000 nœuds en cluster – destiné à tester Hadoop et des modèles analytiques – la manifestation a été l'occasion pour Greenplum d'afficher ses ambitions dans le « *data science* ».



Difficile de trouver meilleure démonstration de l'explosion du Big Data que l'évolution de Greenplum, la division Big Data analytique d'EMC : acquise en juillet 2010 alors qu'elle occupait 175 employés, ils sont désormais 700 à plancher sur le devenir du Big Data.



Scott Yara, cofondateur et vice-président senior produits de Greenplum, nous expose le Big Data comme étant la conjugaison de trois éléments indissociables : une plateforme analytique capable d'optimiser de très gros volumes de données ; la science des données (data science) pour leur trouver de nouveaux usages ; et les applications pour assurer le rendu.

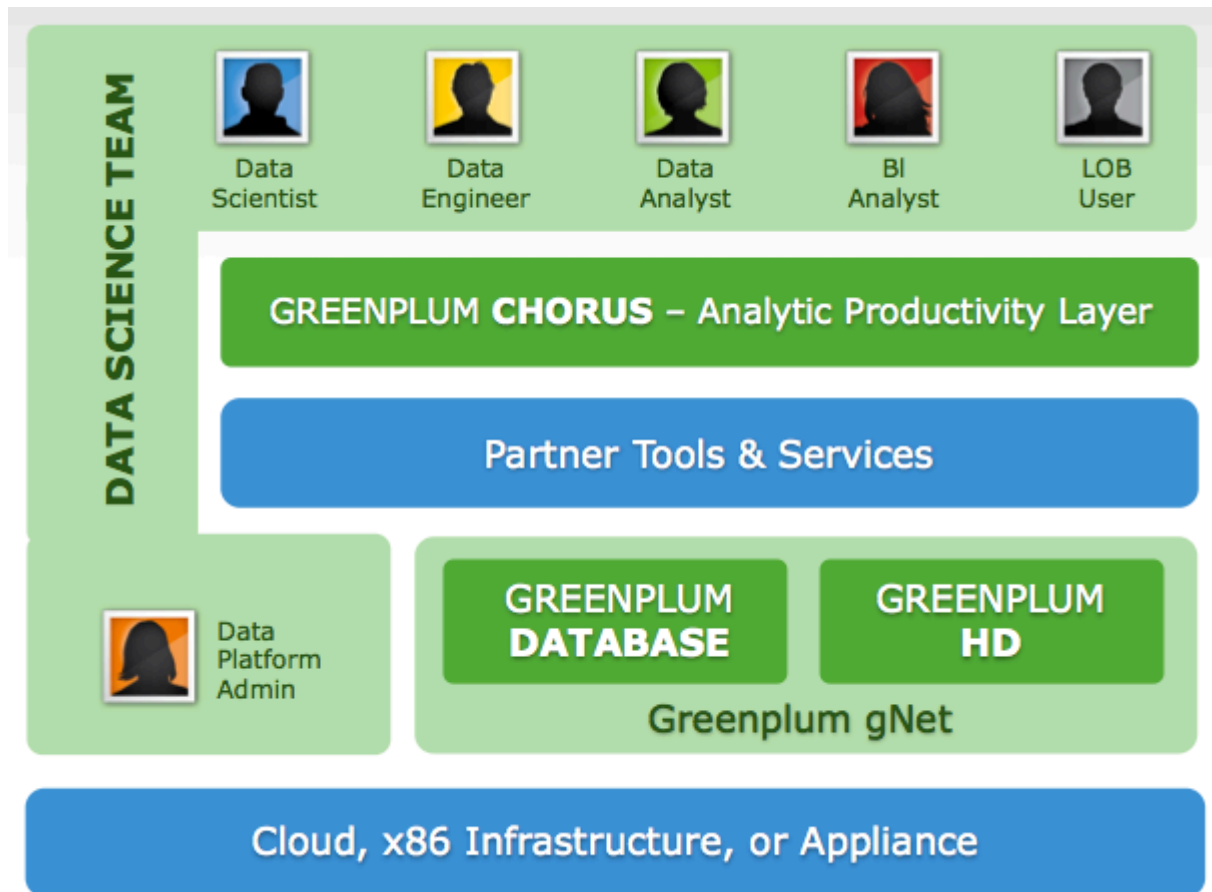
« *Nous devons partir des nouvelles expériences pour les rendre disponibles auprès de nos grands clients* ». Sous-entendu : EMC est capable de construire une plateforme Big Data complète, il suffit de rechercher du côté d'Isilon, dont l'acquisition voici deux ans par EMC prend désormais une nouvelle tournure, celle du scale out, le stockage évolutif, en version Big Data ; d'y ajouter une appliance en rack de data computing qui embarque **Greenplum HD**, la solution Apache-Hadoop prête à l'emploi, la base de données **Greenplum Database** aux processus massivement parallèles, et une couche agile de productivité analytique, **Greenplum Chorus**, qui se dote d'une couleur sociale. L'ensemble logiciel construit la plateforme unifiée d'analyse **Greenplum UAP** (Unified Analytics Platform).

Data Science, la valeur ajoutée du développement

Il manque cependant un élément essentiel à l'édifice, le Data Science. En fait un ensemble de ressources humaines et scientifiques capables d'affronter la difficulté de développer l'analytique Big Data. En la matière, Scott Yara n'y va pas par quatre chemins : la Business Intelligence (BI) est arriérée, elle vient du passé et apporterait peu de valeur, ne fournissant qu'un reporting standard. Le Data Science représente au contraire l'avenir, apportant plus de valeur, en particulier dans sa

capacité à être prédictif. « *Le Big Data réclame de nouvelles compétences et des services.* »

Voilà qui éclaire d'un jour nouveau la récente acquisition de Pivotal Labs par EMC. La start-up édite une solution de gestion de projets de développement et de collaboration temps réel. Il faut bien cela pour accompagner les projets Big Data, et en particulier l'analytique demandée par les grands clients d'EMC. « *Nous proposons une nouvelle approche de l'analytique en mettant à disposition les compétences d'EMC afin de donner à chacun l'opportunité d'accéder à un PhD (doctorat, NDLR).* »



EMC construit autour de Greenplum

EMC crée autour de Greenplum un écosystème de partenaires venant de l'analytique, de la BI, de l'intégration de données, voire de l'industrie. La solution est également testée par des clients, dans la finance, les médias, les technologies, la santé ou des télécoms.

Dans les prochains mois, Chorus sera rendu open source (second semestre) et plusieurs nouveaux produits sont déjà annoncés : **Greenplum DB 4.2**, **Greenplum Command Center** et une configuration Cisco UCS de référence pour Greenplum. Et plus tard l'intégration d'une couche SQL dans les processus Greenplum.