

EMC marie Isilon et Hadoop : le Big Data a-t-il trouvé son maitre ?

« EMC annonce une plateforme pour stocker et analyser ». C'est en ces termes laconiques que **Brian Cox**, Senior Director of Product Marketing de EMC Isilon, nous a présenté en toute confidentialité la première et unique solution de stockage *scale-out* intégrant nativement Hadoop.

Avec le rachat d'Isilon en novembre 2010, EMC s'est offert un acteur innovant du stockage dont la plateforme évolutive (*scale-out*) unique, basée sur une architecture en clusters et bénéficiant d'une vision unifiée via un OS spécifique (OneFS), supporte les protocoles standards de l'industrie (NFS, CIFS, FTP, HTTP et HDFS).

Pendant ce temps, le projet Hadoop de système de fichier open source pour les très gros volumes de stockage de données non structurées a fait son chemin. Il associe le système de fichiers HDFS (*Hadoop Distributed File System*) à Map Reduce, un framework logiciel pour développer des applications supportant de très gros volumes de données parallélisés sur de larges clusters pouvant réunir des milliers de noeuds. Cette association permet en particulier de développer des applications analytiques, c'est à dire d'offrir une solution de Big Data.

Une annonce majeure dans le Big Data

« Les grandes entreprises recherchent une infrastructure de stockage qui offre l'analytique au même niveau. Mais la structure traditionnelle ne permet pas d'associer les deux ! », constate Brian Cox. « Nous savons charger et analyser des données structurées, mais pas les données non structurées. C'est là qu'interviennent le file system Hadoop et Map Reduce pour le processing. En y associant l'avantage de la plateforme *scale-out* d'Isilon, le stockage devient le coeur de l'IT. »

Dans une approche Big Data, l'infrastructure unique proposée par Isilon présente nombre d'avantages, comme le support de très grands volumes de données, la performance, la simplicité dans la gestion, la capacité de faire grossir des portions du système par simple '*plug-and-play*', ainsi que la maîtrise d'une enveloppe budgétaire. Avec son annonce, EMC pousse ce raisonnement plus loin.

Hadoop et l'infrastructure Isilon

« Traditionnellement dans une configuration Hadoop, le stockage et les serveurs, le staff et les équipements sont dispersés », analyse Brian Cox. « Comme Hadoop est déployé sur des grappes de clusters, le repository est présent sur un noeud, mais si celui-ci tombe... C'est le risque pour Hadoop sur ce type d'infrastructure traditionnelle. Hadoop est performant mais affiche des limitations ! Face à ces infrastructures rigides, Isilon apporte la flexibilité, sans downtime. »

En plaçant Hadoop sur le cluster Isilon, le risque que le repository tombe n'a plus de raison d'être, puisque si un noeud tombe les autres prennent le relais. De plus, Hadoop profite de grandes capacités faciles à gérer, sans les limites internes imposées par le stockage traditionnel. Sans oublier

que l'architecture Isilon offre un très haut niveau de protection des données.

Assurer la qualité d'Hadoop

EMC a basé l'offre Big Data d'Isilon sur la distribution Greenplum à laquelle il participe activement. Greenplum adhère au projet Hadoop, qui rappelons-le est open source, hébergé par la fondation Apache. Ce que nous confirme **William Davis**, Product Marketing Manager de EMC Greenplum : « *Les développements d'EMC Isilon sont basés sur le code standard de la distribution Hadoop de la fondation Apache, dont nous suivons les mises à jours. La solution est testée en interne chez EMC, sur une centaine de noeuds, ainsi que déployée chez des clients pour codifier.* »

A qui s'adresse EMC Isilon Hadoop ?

« *Aux grandes entreprises, pas aux sociétés internet comme Google ou Facebook, car elles disposent de leur armée d'ingénieurs qui travaillent à développer leurs propres solutions* », affirme **Brian Cox**. « *L'approche d'EMC sur le Big Data, avec Isilon, Greenplum et les services EMC, est celle du vendeur unique* ». **Quant à savoir si le Big Data va remplacer le datawarehouse ?** « *Pas remplacer mais augmenter l'accès à l'analyse de gros volumes de données. Le datawarehouse restera centré sur les données structurées. Le Big Data ouvre un accès aux nouvelles formes de données.* »