

Tribune : pourquoi Intel lance-t-il une distribution Hadoop ?

Une distribution Hadoop signée Intel... pourquoi ? Notre journaliste **Yves Grandmontagne**, auteur de l'article « [Le Big Data s'enrichit d'une nouvelle distribution Hadoop signée Intel](#) », tente de répondre à cette question.

Même si la réponse peut être multiple, avec sa distribution le géant des processeurs cherche d'abord à avancer ses technologies dans le monde du Big Data. Rappelons simplement que derrière chaque donnée et chaque calcul, il y a une infrastructure avec des serveurs.

Si une certaine tendance est à la parallélisation des flux au travers d'architectures à forte granularité des ressources processeurs, d'où les projets de serveurs équipés de processeurs ARM, le Big Data et son analytique Hadoop sont à l'inverse de très gros consommateurs de ressources de traitement de la donnée et de calcul, donc de processeurs puissants.

Certes, Intel dispose d'un riche portefeuille applicatif, d'une forte équipe de développement et d'un bras armé financier qui investit dans des start-ups avec Intel Capital. L'ensemble contribue à de nombreux projets, souvent open source.

Mais les contributions d'Intel à Apache Hadoop, évidemment reprises dans sa distribution, concernent en priorité le support de la technologie de processeur Intel Xeon. L'optimisation de sa distribution apportée par Intel dans le support de ses disques SSD, en dit également très long sur sa stratégie.

Intel Hadoop vs les serveurs *low-performance*

Un autre motif milite pour l'engagement d'Intel sur Hadoop : les fabricants de serveurs d'entrée de gamme, qualifiés de *low-performance*, qui alignent des processeurs d'entrée de gamme et de basse consommation, pourraient bien attirer les concepteurs de clusters Hadoop, radins comme sont leurs clients ! Ce n'est certainement pas ce que souhaite Intel, dont les marges font de nombreux envieux...

Il suffit de regarder où vont les contributions d'Intel à Hadoop. Elles portent en effet sur HDFS le système de fichiers, YARN le framework de distribution des processus, Hive le module de requêtes SQL, et HBase la clusterisation de la base de données. Tous domaines qui tirent profit de l'optimisation du support par le processeur.

Regarder au-delà d'une stack Hadoop

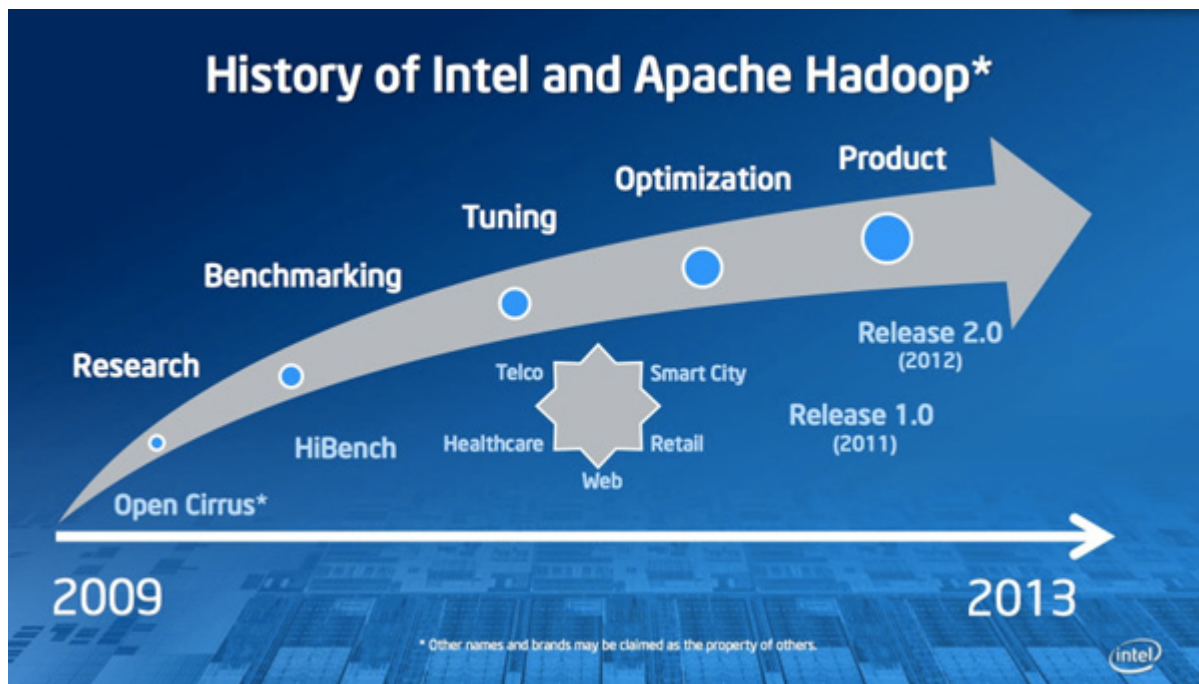
Avec sa distribution Hadoop, Intel vient rappeler qu'il est un acteur incontournable des grandes tendances technologiques des infrastructures IT. Mais il faut regarder au-delà d'une stack logicielle pour mesurer la contribution de ses équipes de développement. Et là c'est bien la mise en avant (ou la défense selon la vision que l'on peut en avoir !) des technologies des processeurs pour

serveurs Xeon qu'il faut voir ici.

Par ailleurs, le domaine du stockage est aujourd'hui largement dominant dans les préoccupations des DSI. L'augmentation exponentielle des volumes de données et la nécessité de les traiter dans une vision analytique porteuse de valeur pour l'entreprise et ses équipes sont une nouvelle contrainte, mais aussi une source potentielle de forts revenus pour tous les fabricants de serveurs. Et de concurrence également, et cela Intel ne l'oublie pas !

Une stratégie sur le long terme

Pour autant, cet engagement n'est pas nouveau, et Intel contribue déjà largement au succès du traitement analytique de très larges volumes de données, comme le rappelle le schéma ci-dessous.



Si les processeurs Intel Xeon sont particulièrement présents tant sur le stockage que sur le Big Data, qui n'en est pas éloigné, peut-être ne le sont-ils pas suffisamment pour le fondeur.

Au final, une infrastructure optimisée pour le Big Data analytique selon Intel devrait être composée de serveurs embarquant des processeurs Intel Xeon, de SSD Intel, de connexions Ethernet 10GbE, et d'Intel Distribution avec ses contributions open source d'optimisation de l'architecture Intel Xeon et SSD, et applicatifs Intel propriétaires...

Comment cela va-t-il se terminer ?

À y regarder de plus près, il n'est pas certain qu'une contribution à la communauté du Big Data analytique comme Intel Distribution for Apache Hadoop, qui en réalité marie allègrement code open source et environnements propriétaires, sur une infrastructure physique qui confond standard et ouverture, viennent réellement remettre les pendules à l'heure dans un domaine qui tend à se complexifier avec la multiplication des offres, toutes ouvertes, mais si restrictives autour de modules au code propriétaire, que l'on peut légitimement se demander comment cela va finir.

Avec sa distribution, Intel ne ferait en réalité que jeter de l'huile sur le feu en proposant un nouveau framework, un de plus ! Certes la firme ne fait pas plus que les autres, mais également pas moins... En revanche, c'est bien sa place dans le datacenter qu'Intel défend, on ne peut le lui reprocher, sauf de se retrancher derrière des arguments communautaires qui en réalité ne feront que peu avancer la machine Hadoop. À moins d'être en ordre de marche avec la stratégie d'Intel !

Voir aussi

[Silicon.fr étend son site dédié à l'emploi IT](#)

[Silicon.fr en direct sur les smartphones et tablettes](#)