

Big Data : Linux, HDFS et SQL s'invitent sur Azure Data Lake

Microsoft renforce aujourd'hui son arsenal dans le secteur de l'analytique Big Data en mode Cloud, via l'offre **Azure Data Lake**. La firme de Redmond y ajoute tout d'abord la mouture définitive de HDInsight en version Linux.

HDInsight permet de créer des **clusters Hadoop** sur Azure Data Lake. Une technologie qui est dorénavant accessible sous Windows Server et sous **Ubuntu Linux**. Une même pile Hadoop est employée, que les utilisateurs aient opté pour un cluster Windows ou Linux. Spark est bien entendu de la partie sur les clusters Hadoop.

« Les clusters HDInsight Linux peuvent être gérés par **Apache Ambari**. Ambari offre la possibilité de personnaliser les paramètres de configuration de votre cluster Hadoop tout en vous donnant une vue unifiée de la performance et de l'état de votre cluster, et en proposant des services de surveillance et d'alerte au sein du cluster HDInsight », [explique Scott Guthrie](#), vice-président exécutif Cloud and Enterprise chez Microsoft.

Les fonctionnalités sont communes aux offres Windows et Linux : élasticité, support des réseaux virtuels, possibilité de scripter les nœuds, etc. Sous Linux, les utilisateurs ne seront pas dépaysés : l'accès **SSH** est proposé et les scripts utiliseront **Bash**. Microsoft signale également qu'il est possible de créer des bases de données NoSQL **HBase** ou **Storm** sur les clusters Linux.

Un dépôt HDFS

Microsoft livre aussi la version de test privée **d'Azure Data Lake Store**, un système de stockage compatible **HDFS**, qui permettra de gérer de larges volumes de données. Cette solution pourra aussi être connectée à un annuaire **Azure Active Directory**. Capable de monter jusqu'à des exaoctets de données, un référentiel de données Data Lake Store se veut aussi bien adapté aux *workloads* classiques qu'aux technologies émergentes, comme **l'Internet des Objets**.

Dernière annonce de la journée, la présentation (en version de test privée là aussi) **d'Azure Data Lake Analytics**, un service Cloud d'analytique Big Data, qui se veut simple d'utilisation et pourra interagir avec des sources de données Azure Data Lake Store ou Azure SQL Database. Cette solution s'appuie **U-SQL**, une technologie adaptée spécifiquement au Big Data. Les utilisateurs pourront mettre à profit leurs connaissances de MS-SQL et de la plate-forme .NET, qui sont reprises par U-SQL. « U-SQL vous permet de définir des tâches Big Data, et de facilement y inclure votre propre code », résume Scott Guthrie. Une offre utilisée précédemment en interne par la firme sur des jeux de données de plusieurs exaoctets.

Microsoft livre des outils permettant de **déboguer et optimiser** des tâches U-SQL depuis Visual Studio. Un élément essentiel, puisqu'une requête bien optimisée consommera moins de ressources et coûtera donc moins cher lors de son lancement sur le Cloud Azure.

À lire aussi :

[GS-series : des VM taille maousse pour le Cloud Azure](#)

[Azure Data Lake, du stockage Hadoop sur le Cloud de Microsoft](#)

[Microsoft multiplie par 32 la capacité de stockage d'Azure Backup](#)