

Apache vient booster Hadoop avec Spark

1.0

Présenté comme le couteau suisse de la pile Big Data Hadoop, [Spark](#) est aujourd'hui livré en version 1.0 par la Fondation Apache.

Spark est un outil dédié aux traitements de données à large échelle, qui propose de **multiples innovations** :

- Il affiche des vitesses de fonctionnement très avantageuses par rapport à **MapReduce** de Hadoop : jusqu'à x10 sur disque et jusqu'à x100 en mémoire.
- Les applications Spark pourront être écrites en divers langages de programmation : **Java, Scala ou Python**.
- Spark peut s'intégrer à un cluster Hadoop ou fonctionner de manière autonome et est capable de lire des données depuis **HDFS, Hbase ou Cassandra**.

Au besoin, Spark est même capable de se connecter à des bases SQL, ce qui en fait un outil particulièrement flexible, à la croisée des chemins entre les bases de données classiques et les solutions dédiées au Big Data.

Une adoption très rapide

Spark est tout indiqué pour créer des applications d'analyse de données sur des jeux massifs d'informations. Il a d'ores et déjà été adopté par certains grands noms, comme Alibaba, la Nasa ou encore Yahoo. La Nasa compte par exemple l'employer dans le cadre de la mise au point de modèles climatiques.

Notez que Spark est aujourd'hui présent dans la plupart des distributions Hadoop. Les offres de **Cloudera, IBM, Intel, MapR et Pivotal** sont ainsi citées.

Voir aussi

[Quiz Silicon.fr – Testez vos connaissances sur le Big Data](#)