

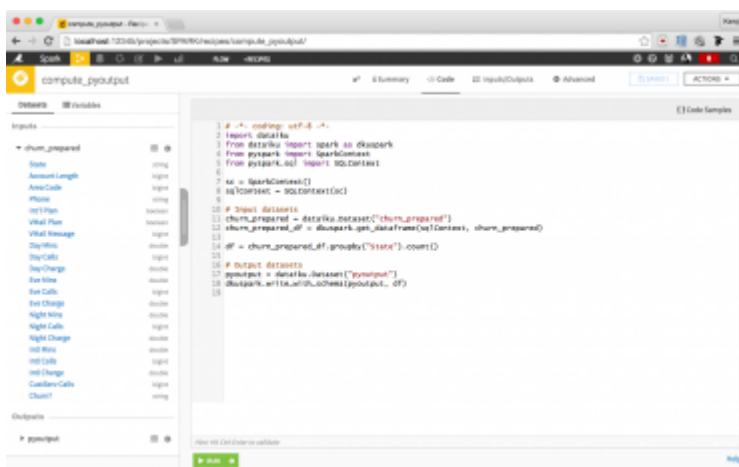
Big Data : Data Science Studio joue aussi la carte Spark

C'est fait ! **Spark** complète la version 2.1 de [Data Science Studio \(DSS\)](#), plateforme logicielle d'analyse de données et de création d'applications prédictives conçue par l'éditeur français **Dataiku**. Spark est un framework de calcul distribué. Il utilise des techniques In-Memory et peut fonctionner sur un cluster de serveurs. C'est un projet Open Source phare de la Fondation Apache désormais. « *Spark avec sa release 1.5 atteint le niveau de maturité permettant un déploiement efficace pour tous, et rejoint donc, avec Cassandra, MongoDB et Hadoop MapReduce, la galaxie des technologies disponibles dans DSS* », explique à la rédaction **Florian Douetteau**, cofondateur et CEO de Dataiku. Avec Spark, DSS peut traiter des volumes de données plus importants qu'avec un moteur de traitement classique.

DSS s'appuie sur le duo Hadoop + Spark

Les frameworks, les langages de programmation et les dialectes (Py, R, Pig, Hive, SQL, Spark...) évoluant très rapidement, « *nous voulons fournir à nos clients le meilleur des technologies* » via une interface unifiée, précise le cofondateur de Dataiku. L'utilisateur doit pouvoir « *passer rapidement de l'une à l'autre, sans devoir développer à nouveau toute son application* », ajoute-t-il. **MapReduce** est donc toujours disponible dans Data Science Studio. « *À vrai dire, pour certaines catégories d'usages, MapReduce est plus efficace que Spark actuellement* », souligne Florian Douetteau. DSS supporte aussi Apache **Tez** (une amélioration de MapReduce poussée par Hortonworks) et **Impala** (moteur de requêtes initialement développé par Cloudera, mais disponible dans la plupart des distributions Hadoop).

Résultat, DSS peut être l'outil principal de développement de projets, de la détection de fraude à la réduction du churn, en passant par la maintenance prédictive.



Dataiku s'exporte outre-Atlantique

Créée en 2013, la start-up Dataiku a été profitable dès ses débuts. Depuis janvier 2015 son équipe a doublé, passant de 25 à 50 collaborateurs répartis [entre Paris et New York City](#). L'entreprise compte aujourd'hui plus de 60 entreprises clientes, dont « *quelques grands comptes aux États-Unis* ».

Lire aussi :

[F. Douetteau, Dataiku : « Nul besoin de data scientists superstars pour le Big Data
IBM investit massivement dans la technologie Big Data Spark](#)

crédit photo © Dataiku