

# Google veut faire de Dataflow un projet incubateur Apache

Dataflow est le successeur de MapReduce, FlumeJava et Millwheel, programmes auxquels les chercheurs de Google ont largement contribué. Dataflow, technologie de traitement de données en mode batch ou temps réel, inclut un modèle unifié de programmation, un kit de développement ([DataflowJavaSDK](#) déjà sous licence Apache) et des *runners* pour exécuter les pipelines de flux de données dans des frameworks (Apache Spark, Apache Flink ou Google Cloud Dataflow).

Pour accélérer ses développements, étendre la communauté de contributeurs et renforcer les liens avec d'autres projets Open Source, Google [propose de transférer Dataflow](#) à la Fondation Apache, afin d'en faire un projet incubateur Apache (qui filtre les projets Open Source que l'ASF peut porter).

« Cette proposition est une étape dans la mise en œuvre d'une architecture de données qui répond à des besoins multiples de traitement (et) fonctionne sur site, dans le Cloud ou localement. Google Cloud Dataflow restera un service managé 'no-ops' pour exécuter rapidement et au meilleur coût des pipelines Dataflow dans Google Cloud Platform », ont expliqué dans [un billet](#) de blog Frances Perry et James Malone, respectivement développeuse et manager produit chez Google. Le projet incubateur souhaité par Google est soutenu par d'autres acteurs, dont Cloudera, data Artisans, Talend, Cask et PayPal.

**Lire aussi :**

[Google I/O : Dataflow alimente en temps réel Big Query en Big Data](#)

**crédit photo © rvlsoft / Shutterstock.com**