

Open Source : Delta Lake porté par la Fondation Linux

En avril 2019, Databricks, l'entreprise fondée par les créateurs du moteur de traitement data [Apache Spark](#), a déclaré [rendre open](#) source [Delta Lake](#).

Six mois plus tard, la [Fondation Linux](#) annonce héberger le projet Delta Lake. Une strate de stockage lancée en 2017, conçue pour garantir l'intégrité de données brutes et massives stockées dans de vastes référentiels (ou lacs) de données. Les « data lakes ». Le projet fédère aujourd'hui plus de 4000 organisations, parmi lesquelles les groupes Intel, Alibaba et Booz Allen Hamilton (ancien employeur du lanceur d'alerte américain [Edward Snowden](#)).

Plus de deux exaoctets de données sont traitées chaque mois, selon ses promoteurs.

Gouvernance ouverte

« Chaque entreprise veut obtenir davantage de valeur des données grâce à la data science, au machine learning et à l'analyse de données », [a souligné](#) la Fondation Linux. Mais les données non structurées ou semi-structurées peuvent être compromises.

C'est ici que Delta Lake intervient.

« Delta Lake résout les problèmes de fiabilité des données en rendant les transactions compatibles ACID (atomicité, cohérence, isolation et durabilité). [Pour] garantir que le lac de données est exempt de données corrompues et non conformes. »

En outre, le modèle « ouvert » de gouvernance du projet promu doit encourager la participation et la contribution technique bien au-delà de l'écosystème Apache Spark.

À la fois sur site (on-premise) et dans le cloud.