



BIG

Algorithmes PySpark pour traitement de données

DATA

Nous créons environ 2,5 trillions d’octets de données par jour. En 2 ans nous avons créé 90% des données présentes dans le monde. Ce nombre de données provient de l’avancée de toutes les technologies et aussi de nos habitudes usuelles telles que l’envoi de messages, mails ou vidéos. C’est pourquoi il en devient difficile de travailler sur ce volume de données avec les outils classiques de gestion de base de données. C’est pourquoi la manipulation des données représente bel et bien un enjeu extrêmement important pour notre futur.