

# Pourquoi le moment est-il venu d'alléger vos **entrepôts de données** ?

Depuis des décennies, les entreprises utilisent des entrepôts de données à des fins pour lesquelles ils n'ont pas été conçus, notamment les processus ETL (Extraction-Transformation-Loading) et le stockage d'énormes volumes de données inutiles.

Dans le même temps, la nature des données a changé, tout comme les pratiques d'analyse, ainsi que les méthodes de stockage et d'accès aux données, devenues à la fois plus efficaces et économiques.

Pour autant, les entrepôts de données restent des outils efficaces pour traiter des données propres, intégrées et fiables, utilisées pour produire des analyses comme des rapports, des comparaisons historiques, des analyses de client et des calculs d'indicateurs de performance clés (KPI). Cependant, la plupart des entrepôts de données d'entreprise sont dépassés pour plusieurs raisons :

## 1. Ils sont trop coûteux pour mettre en œuvre des programmes ETL

- Les processus ETL peuvent consommer 50 % et plus des capacités de traitement des entrepôts de données.
- Plus vous consacrez un entrepôt de données à un processus ETL, moins vous disposez de capacités pour les applications d'aide à la décision.

## 2. Ils se transforment en maquis de données

- Près de 50 % des données contenues dans un entrepôt de données d'entreprise ne sont jamais utilisées.<sup>1</sup>

## 3. Ils ne permettent pas d'incorporer des sources de données non traditionnelles

- Les flux de données issus des médias sociaux, des parcours de navigation et des capteurs de l'Internet des Objets exigent d'autres structures que les tableaux traditionnels, organisés en lignes et colonnes.

## 4. Ils ne permettent pas l'exploration des données

- Une modélisation lourde des données est nécessaire pour permettre l'exploration, ce qui limite leur utilisation, par exemple dans les processus de découverte des données.

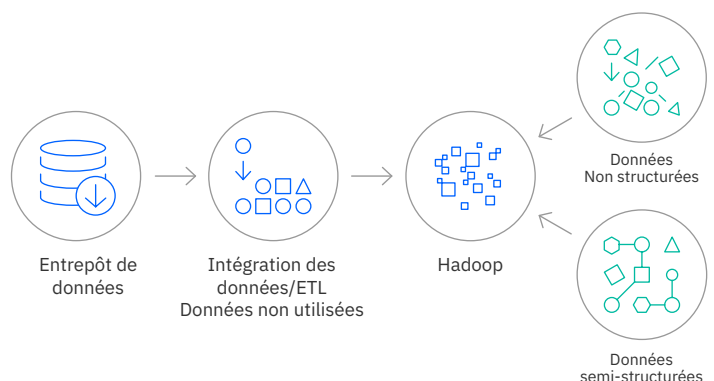
## Moderniser les outils d'analyse d'entreprise

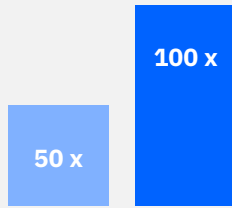
Les infrastructures — qu'il s'agisse de routes, de ponts ou de systèmes informatiques — doivent être modernisées pour répondre à des nouveaux besoins. Dans le cas d'un entrepôt de données, moderniser l'architecture d'analyse nécessite une mise à niveau et une transformation profonde des traitements (accès, stockage, préparation, gouvernance, analyse).

L'une des approches les plus efficaces à cet effet consiste à transférer les données des entrepôts et les traitements ETL vers une plate-forme Apache Hadoop. Il est ainsi possible de réduire les coûts tout en gérant efficacement l'intégration, la qualité et la gouvernance des données.

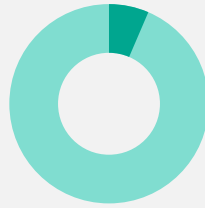
## Trois activités distinctes et interdépendantes sont nécessaires pour ce délestage

À chaque phase de ce processus de transfert, l'entreprise peut introduire des informations de bonne qualité dans un « lac de données » ou un référentiel Hadoop pour améliorer la qualité des données et contribuer à intégrer les connaissances acquises dans les opérations futures de l'entreprise.

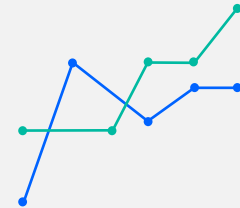




**50 x – 100 x**  
réduction des coûts de traitement et de stockage<sup>2</sup>



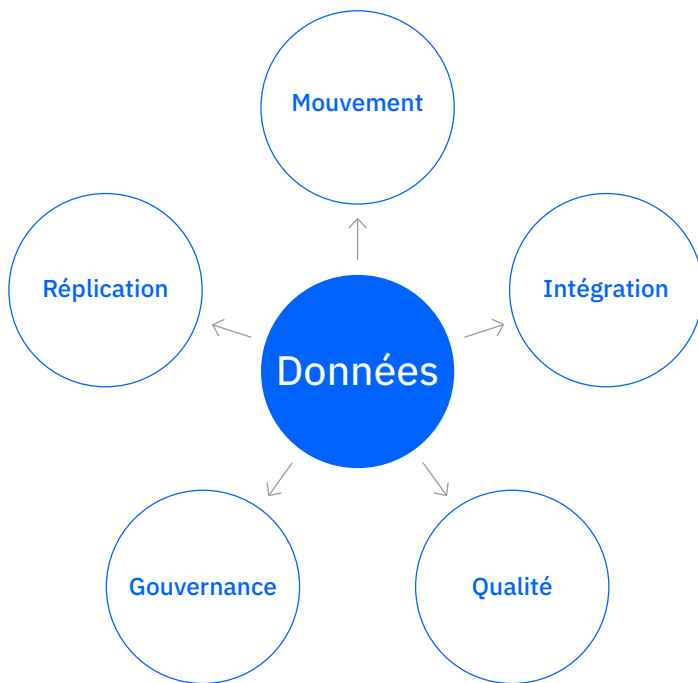
**10 x**  
gain de productivité grâce aux outils vs codage manuel<sup>3</sup>



**Analyse**  
enrichie des données non traditionnelle<sup>3</sup>

### Cinq fonctionnalités pour un délestage efficace

Toute entreprise se doit de considérer cinq capacités pour créer une stratégie de délestage permettant de maximiser le retour sur investissement (ROI) de l'opération.



**Mouvement** : Extraire, transférer et assimiler des volumes de données importants sans stockage sur disque, ni nouveau codage.

**Réplication** : Assurer un transfert des données à faible latence entre les systèmes sources et les cibles Hadoop pour optimiser l'utilisation des ressources tout en assurant la mise à jour au bon moment.

**Gouvernance** : Assurer la gestion des données grâce à des intervenants capables de localiser et d'extraire des informations à propos d'objets de données, leur signification, leur localisation physique, leurs caractéristiques et leur utilisation grâce à des politiques de gouvernance adoptées à l'échelle de l'entreprise et à un « vocabulaire métier ».

**Intégration** : Mettre en œuvre des centaines de processus d'intégration de données, créés une fois pour toutes et appliqués partout, sans modification.

**Qualité** : Définir des règles de validation appliquées de manière cohérente sur d'innombrables sources de données ; appliquer les règles « à la volée » pour ne charger, dans les systèmes cible, que des données validées ; surveiller et évaluer les sources de données pour suivre et garantir la conformité.

#### Lire le blog [➔](#)

Découvrez trois idées pour vous engager dans la démarche de plate-forme de données unifiée.

#### Participer au webinar [➔](#)

Découvrez comment les clients qui ont modernisé leur architecture d'analyse ont réussi à résoudre les défis des entrepôts de données d'aujourd'hui grâce aux possibilités de délestage.



© Copyright IBM Corporation 2018. IBM, le logo IBM et [ibm.com](http://ibm.com) sont des marques d'International Business Machines Corp., déposées dans de nombreux pays du monde. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. Une liste actualisée des marques déposées IBM est accessible sur le Web sous la mention « Copyright and trademark information » à l'adresse [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml). Ce document est considéré comme à jour à sa date initiale de publication et peut être modifié par IBM à tout moment. Toutes les offres ne sont pas disponibles dans chaque pays dans lequel opère IBM.

LES INFORMATIONS CONTENUES DANS CE DOCUMENT SONT FOURNIES « EN L'ÉTAT », SANS AUCUNE GARANTIE EXPRESSE OU TACITE, NOTAMMENT SANS AUCUNE GARANTIE DE QUALITÉ MARCHANDE OU D'ADAPTATION À UN EMPLOI SPÉCIFIQUE, ET SANS AUCUNE GARANTIE OU CONDITION DE NON INFRACTION VIS-À-VIS DES LOIS. Les produits IBM bénéficient d'une garantie, conformément aux conditions générales des contrats dans le cadre desquels ils sont fournis.

1 Forrester : The Next-Generation EDW Is The Big Data Warehouse (2016)  
 2 ITG : Business Case for Enterprise Data Integration Strategy: Comparing IBM InfoSphere Information Server and Open Source Tools (2013)  
 3 ITG : Business Case for Enterprise Data Integration Strategy: Comparing IBM InfoSphere Information Server and Open Source Tools (2013)