



Construire un Data Lake dans le Cloud

OBTENIR UNE SOURCE UNIQUE DE DONNÉES
FIABLE RAPIDE ET SÉCURISÉE



Sommaire:

- 3 Le data lake : mythes et réalité
- 5 Les composantes de votre data lake
- 7 Les facteurs clés de succès
- 8 Étape 1 : Unifier des sources de données multiples
- 10 Étape 2 : Sécuriser les données par la gouvernance
- 11 Étape 3 : Mettre ses données en qualité
- 13 Étape 4 : Faciliter le libre-service
- 15 Tenir les promesses de l'analyse prédictive
- 16 Pour en savoir plus

Le data lake : mythes et réalité

Les organisations qui ont une approche orientée données cherchent à tirer des enseignements à partir de toutes leurs données au travers de leurs activités et des exigences de leurs clients. Elles collectent et analysent des volumes toujours plus importants de données issues de sources traditionnelles, telles que les ERP, les CRM et les systèmes de points de vente, mais aussi de sources nouvelles, à savoir les logs de connexion, les applications web et mobiles, les objets connectés (IoT) et d'autres encore.

Cependant, cette démarche nécessite de disposer d'un référentiel unique pour stocker, réconcilier et exploiter facilement et efficacement les données. Mais charger un tel volume d'informations dans un data warehouse traditionnel peut se révéler compliqué, long, coûteux voire impossible. Les données provenant de nouvelles sources arrivent souvent en formats semi-structurés et nécessitent d'être transformées et traitées avant le chargement. De plus, le coût et la complexité des opérations liées à la conservation de vastes quantités de données brutes dans un entrepôt classique à partir d'un nombre toujours croissant de sources seraient excessifs.

Le data lake a fait son apparition il y a plus d'une décennie dans le but de résoudre les problématiques suivantes :

- Créer un référentiel de données évolutif et peu onéreux pour rassembler les données brutes issues de sources multiples, puis les découvrir et les croiser.
- Rendre ces données affinées, organisées et certifiées accessibles au plus grand nombre vers d'autres systèmes, y compris un data warehouse, pour effectuer des tâches d'analyse, de découverte de données en libre-service et de reporting avancées.

Une idée a priori simple mais qui dans la pratique ne l'est pas forcément. Sans la bonne technologie, une qualité de données acceptable et une gouvernance appropriée, le « lac » peut bien vite se transformer en marécage – un réservoir de données difficiles à utiliser et à comprendre, et dans la pratique inaccessible à ceux qui en ont vraiment besoin. Plus les données qui se déversent dans le data lake sont nombreuses et diversifiées, plus le problème prend de l'ampleur. Il est alors difficile d'en extraire des informations significatives et donc d'en tirer de la valeur.





VERS UNE MEILLEURE SOLUTION

Réussir la construction d'un data lake demande, dès le début, une conception capable d'incorporer la data stewardship, la gouvernance et la sécurité, tout en permettant un accès facile aux données. On part souvent de l'hypothèse qu'un data lake nécessite plusieurs technologies déployées en tant que solutions indépendantes.

Toutefois, les progrès dans les domaines du data warehouse et de la gestion des données permettent maintenant de concrétiser les promesses initiales du data lake, sans l'inconvénient de technologies hétérogènes et limitées qui entraînent de nouveaux silos de données.

Construire un data lake moderne nécessite une architecture permettant de stocker aisément des données brutes, de les explorer immédiatement, de les affiner de manière cohérente et contrôlée, et de traiter efficacement un large éventail de rapports et d'analyses.

Pourquoi un data lake moderne

De manière générale, la nécessité d'implémenter un data lake moderne s'est accélérée avec la convergence de trois tendances importantes :

- Les applications SaaS sont de plus en plus nombreuses. En parallèle, les données sont de plus en plus traitées dans le Cloud.
- Des volumes exponentiels de données arrivent à tout moment, ce qui implique la prise de décisions en temps réel, avant que les données ne deviennent obsolètes ;
- L'accès aux données, traditionnellement contrôlé par les SI, migre vers le libre-service pour les utilisateurs métier.

Les services informatiques subissent de plein fouet la puissance de ce changement profond et rapide. En réalité, le déploiement de data lakes au sein des organisations n'a fait qu'ajouter des tâches supplémentaires à l'emploi du temps déjà surchargé du département IT. Ce type de projet ne dispose pas de la flexibilité, de la gouvernance et de la gestion de données nécessaires pour rester en adéquation avec les nouvelles formes d'analyse.

Les composantes clés d'un data lake

LES INGRÉDIENTS CLES POUR DEVENIR UNE ENTREPRISE ORIENTÉE DONNÉES

DES BIG DATA, MAIS PAS À N'IMPORTE QUEL PRIX

Les données d'aujourd'hui sont issues d'une multitude de sources, y compris les bases de données relationnelles et NoSQL, les dispositifs IoT et les données générées par les applications SaaS et d'entreprise. Pour les plateformes traditionnelles, réunir toutes ces informations constitue un véritable défi.

Pour y faire face, des types différents de données sont habituellement conservés sur des plateformes différentes. Cette approche crée des « îlots » de données éloignés, qui augmentent la complexité et cachent des connaissances potentielles.

Un des objectifs principaux d'un data lake est de réunir ces informations. Hélas, de nombreux data lakes sont implémentés en tant qu'« îlots » de données indépendantes, non reliées à d'autres structures, telles que les data warehouses et les data marts, supportant les analyses et le reporting.

Il en résulte que seul un petit groupe d'ingénieurs et de data scientists peut accéder aux données contenues dans le data lake. Les autres utilisateurs doivent attendre que ces spécialistes préparent et exportent les données pour

eux. Cependant, écrire manuellement des scripts pour normaliser les données en un format standardisé et en élargir l'accès au plus grand nombre peut se révéler très chronophage et onéreux. D'autre part, confier le data lake uniquement aux mains expertes des data scientists limite l'usage du datalake et ralentit le fonctionnement de l'entreprise.

GOVERNANCE, SÉCURITÉ ET CONFORMITÉ

Avec l'explosion des données, les utilisateurs à travers les services de l'entreprise sont de plus en plus nombreux à demander un accès aux informations. Par conséquent, des règles solides de gouvernance sont essentielles pour rendre disponibles les données là où c'est nécessaire, et en garantir un bon usage. Par exemple, tout type de données peut contenir des éléments devant être gérés et sauvegardés selon des règles spécifiques. Certaines informations relèvent aussi de réglementations rigoureuses, telles que la législation sanitaire (HIPAA), la norme de sécurité de l'industrie des cartes de paiement (PCI DSS), le règlement général sur la protection des données de l'Union européenne (GDPR). Les données pouvant nécessiter une gouvernance spécifique comprennent :

- les informations relatives aux cartes de crédit ;
- les numéros de sécurité sociale ;
- la date de naissance ;
- l'adresse ;
- les informations sur le réseau IP ;
- les coordonnées de géolocalisation.

Les données au sein d'un data lake doivent se conformer à ces obligations. L'accès aux données et leur utilisation doivent être contrôlés, suivis et sécurisés à tout moment aussi bien lors du stockage dans une plateforme de données que pendant leur chargement et leur intégration. De tels standards nécessitent des outils de gestion et de gouvernance. Le but est de disposer d'un contrôle strict des accès, du cryptage de données au repos et en mouvement, d'enregistrements vérifiables des modifications effectuées.



UNE VUE DES DONNÉES COHÉRENTE ET FIABLE

Disposer de beaucoup de données, oui, mais sans sacrifier la qualité. Les « data scientists » ont besoin d'accéder à des données brutes avant qu'elles soient nettoyées et normalisées, afin de les explorer et conduire des expériences. Le reste de l'entreprise, en revanche, a besoin d'une vue cohérente et fiable des données pour le reporting et les analyses.

Le data lake doit être à la fois la source ultime pour explorer vos données et en même temps satisfaire aux exigences et aux contraintes inhérentes à l'analyse et au reporting. Pour cela, les données doivent être normalisées au bon format, avec des métadonnées vérifiables qui indiquent la provenance des données et le moment du chargement. La normalisation des données et d'autres éléments inhérents à la qualité sont nécessaires pour permettre aux SI de fournir aux utilisateurs métier des fonctionnalités en libre-service, tandis que les métadonnées activent les sous-ensembles de données qui pourront ainsi être facilement localisés, profilés et éliminés (ou nettoyés) si la qualité est insuffisante.

AU SERVICE DE TOUS LES UTILISATEURS

De plus en plus de professionnels souhaitent accéder aux données d'entreprises : qu'ils soient analystes, data scientists, simples utilisateurs ou professionnels de la donnée, leur nombre augmente sans cesse. Dispersés et isolés dans différents services de l'entreprise, tous ont besoin d'un accès facilité aux données.

Universaliser l'accès des données d'un data lake réclame la bonne méthodologie et la technologie appropriée.

Le libre-service est devenu essentiel pour élargir l'accès aux données au sein de l'entreprise. Le généraliser nécessite la combinaison d'une plateforme de données et d'une solution de gestion de données pour faciliter, gérer et sécuriser l'accès aux données. Par exemple, la plateforme de données requise doit contourner les étapes manuelles du déploiement et éviter des réglages laborieux. Elle doit également soutenir les charges de travail imprévues et un nombre important d'utilisateurs simultanés. Sans cela, il sera difficile au département IT de rendre les données accessibles à tous.

Construire un véritable data lake

LES QUATRE PILIERS

1

UNIFIER DES SOURCES
DE DONNÉES MULTIPLES



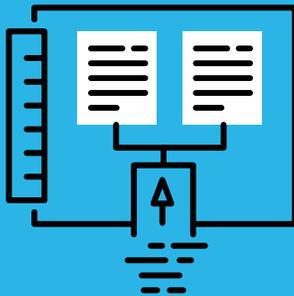
2

SÉCURISER LES DONNÉES
PAR LA GOUVERNANCE



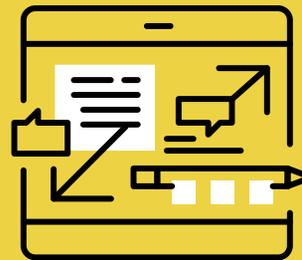
3

METTRE SES DONNÉES
EN QUALITÉ



4

FACILITER LE
LIBRE SERVICE



ÉTAPE

Unifier des sources de données multiples

CANALISEZ TOUTES VOS DONNÉES VERS UN RÉFÉRENTIEL UNIQUE

Si votre entreprise n'est pas encore orientée données, elle va le devenir, si elle s'appuie sur un ensemble de sources différentes de données : NoSQL et data stores SQL relationnels, applications SaaS et traditionnelles, données IoT. Ces sources ont des formats, des modèles et des structures très variés. Vous aurez besoin de plateformes et d'outils à la pointe pour simplifier au possible la gestion et la consolidation de toutes ces données.

UN DATA LAKE PERFORMANT DANS UN ENTREPOT DE DONNÉES CONÇU POUR LE CLOUD (OU « CLOUD DATA WAREHOUSE »)

Faites simple quand vous envisagez le datalake. Pour centraliser toutes vos données en un seul endroit, privilégiez une infrastructure Cloud. Grâce à la puissance du Cloud, vous pouvez en effet envisager un datalake dans un data warehouse. Le cloud data warehouse devient ainsi la solution idéale pour stocker et centraliser nativement des données dispersées à un prix raisonnable.

Le cloud data warehouse doit néanmoins remplir certaines conditions :

- **charger et analyser immédiatement les données brutes**, sans passer par la transformation. Il doit vous permettre d'interroger les données, tout de suite après les avoir chargées dans le data lake ;
- **gérer des données structurées et semi-structurées**. Ces deux types de données devraient coexister de manière harmonieuse. Idéalement, vous devriez être en mesure de créer des tables simples dans le cloud data warehouse et de diffuser des données structurées et semi-structurées, sans codage ni intervention manuelle ;
- **écrire des requêtes SQL tant pour les données structurées que pour les semi-structurées** sans programmation supplémentaire. Un cloud data warehouse devrait utiliser du SQL natif et appliquer une approche « schema on read » aux données semi-structurées ;
- **séparer le traitement et le stockage**, pour conserver des volumes massifs de données brutes au meilleur coût possible, tout en déployant uniquement la capacité de calcul nécessaire.

Le cloud data warehouse devrait être alimenté par un connecteur complètement natif pour maximiser les fonctionnalités de type multitenant supportant ainsi le chargement de données en parallèle. Qu'importe le nombre de tables et de sources différentes : un connecteur optimisé assure une ingestion des données au débit maximum dans le data lake.

CONDITIONS POUR LE CHARGEMENT DES DONNÉES

Les données changent constamment. Au fur et à mesure qu'elles arrivent, elles pourraient nécessiter un traitement analogue aux phases de profilage, de normalisation, d'agrégation et de nettoyage. Accélérer le chargement des données est une étape essentielle lorsqu'on réunit des sources de données variées. Les meilleures pratiques pour créer une plateforme solide pour l'ingestion des données comprennent :

- **la connexion aisée à de sources multiples de données**, sans élaborer des scripts de programmation importants et collecter toutes les données partout où elles se trouvent ;
- **l'ubiquité par lots et en continu** : manipuler les données historiques et le chargement de données en temps réel, traiter les pipelines de données au fur et à mesure de leur entrée ;
- **l'évolutivité des volumes et des variétés de données** : intégrer rapidement de nouvelles sources, telles que les données des internautes, les parcours de navigation Web, les médias sociaux et les dispositifs connectés.

GESTION DU CYCLE DE VIE COMPLET

Pas de vrai data lake sans les bons outils d'intégration : profilage, agrégation, normalisation et nettoyage sont essentiels à chaque étape du cycle de vie de la donnée.





Sécuriser les données par la gouvernance

TIREZ PROFIT DES MÉTADONNÉES ET DE L'ARCHITECTURE DU DATA LAKE

Avec toutes les données qui se déversent dans votre data lake, vous mélangez probablement des données internes et externes à votre entreprise. Or, beaucoup de services demandent à accéder justement à ces données, qui peuvent inclure des informations sensibles.

Face aux cas d'usage multiples, une gouvernance adaptée est nécessaire pour que les utilisateurs disposent de données réactualisées, pertinentes et précises, mais surtout qu'ils accèdent uniquement aux données autorisées. Une solution data lake performante doit permettre une gouvernance efficace :

- en **ajoutant du contexte aux métadonnées** pour fournir des informations telles que l'origine des données, l'auteur et la nature des modifications, ou encore les relations entre les divers jeux de données ;
- en **conservant les données** par le stewardship et la préparation des données, tâches effectuées par des utilisateurs capables de qualifier les données avec précision ;
- en **activant un processus collaboratif de gouvernance de données**. Au lieu d'une gouvernance traditionnelle de type descendante, la solution data lake offre une approche collaborative pour

garantir la fiabilité et la qualité des informations contenues dans le data lake. Votre data lake deviendra ainsi la source unique de données fiables.

FACILITER LA GOUVERNANCE DANS LE CLOUD DATA WAREHOUSE

La mise en place de ces contrôles nécessite le soutien de la plateforme de données sous-jacente. Afin d'être utilisé en tant que data lake, un data warehouse doit :

- assurer le cryptage efficace et la gestion des clés pour toutes les données ;
- fournir et appliquer un contrôle d'accès granulaire configuré par utilisateur et par rôle ;
- tenir à jour un registre des accès effectifs et des tentatives d'accès aux données et au data warehouse.

Les métadonnées générées au sein d'un cloud data warehouse peuvent supporter aussi les contraintes liées à la gouvernance. Par exemple, un data warehouse peut générer des métadonnées à partir des données chargées et les rendre accessibles aux requêtes.

ÉTAPE 3

Mettre ses données en qualité

CONSTRUISEZ UN DATA LAKE AGILE, PRÊT POUR LA PRODUCTION

Pourvoir aux besoins de la Business Intelligence (BI) est souvent l'objectif pour la construction d'un pipeline de données. Mais il est vain de construire une vue analytique sur un datalake mal structuré. Un data lake sans qualité de données rend la généralisation des données en libre-service difficile, voire impossible. Sans nettoyage préalable des données, le data lake ne pourra pas être un support fiable pour la BI. Sans des données fiables, le risque est alors de ne pas prendre les bonnes décisions et donc de nuire à la conduite de l'entreprise.

Un data lake conçu sur un cloud data warehouse gagne en agilité et en rapidité. Il offre un système unique et intégré pour stocker facilement et accéder à de vastes quantités de

données. Dès lors, le datalake devient la source commune où chacun puisera ses données selon son rôle.

Si toutes vos diverses sources de données et métadonnées sont intégrées dans un seul système, vos utilisateurs peuvent accomplir leurs tâches d'analyse sur des données fiables. De plus, les utilisateurs bénéficieront de l'évolutivité et de la flexibilité dont ils ont besoin pour explorer le data lake sans être obligés de déplacer les données vers un système différent.



UNIFIER LA GESTION DES DONNÉES

Pour faire de votre data lake moderne la source unique des données d'entreprise, vous devez disposer d'une infrastructure pour la gestion des données incluant

- des fonctionnalités intégrées de masquage et de **qualité étendue des données**
- **une mise en œuvre opérationnelle cohérente** pour améliorer la qualité et l'agilité des données et l'agilité ;
- **une plateforme unique pour tous les cas d'utilisation et tous les profils utilisateurs**, afin d'augmenter la productivité et la collaboration entre les équipes.

L'association d'un cloud data warehouse et d'un logiciel pour la gestion des données fournit une solution adaptée. Les données résident dans une plateforme unifiée qui prend en charge et simplifie l'administration et la gouvernance. La donnée est gérée de bout en bout et en toute cohérence. Elle est prise en charge par **une plateforme de données unifiée qui gère et simplifie la gestion des données et sa gouvernance.**

UNE MEILLEURE APPROCHE À LA TRANSFORMATION DES DONNÉES

Avec cette nouvelle approche, vous pouvez personnaliser *vos transformations de données*. La plateforme de gestion de données destinée au data lake devrait offrir la possibilité de développer des transformations personnalisées. Cela tranche complètement avec l'approche *classique d'intégration de données* qui consiste à restreindre les possibilités à une *bibliothèque de transformations préconfigurées et donc limitées*.

Face à des schémas de données en constante évolution, les approches classiques ne fonctionnent plus car elles n'ont pas la souplesse nécessaire pour les supporter.

Un cloud data warehouse quant à lui reste complètement évolutive pour transformer les données ; son coût de stockage reste faible puisqu'il ne repose pas sur l'informatique de l'entreprise. Cela vous permet d'exploiter pleinement son évolutivité et sa puissance, et par là même d'accélérer la transformation des données.

ÉTAPE 4

Faciliter le libre-service sur vos données.

DONNER ACCÈS À DES DONNÉES GOUVERNÉES EN VOUS APPUYANT SUR UNE INFRASTRUCTURE DE STOCKAGE ET DE TRAITEMENT ÉVOLUTIVE

Votre data lake n'atteindra son plein potentiel que si vous élargissez l'accès aux données au plus grand nombre d'utilisateurs. Mais pour cela, il faut trouver un équilibre. En effet, votre objectif n'est pas nécessairement de permettre à tous d'accéder à n'importe quelle donnée en libre-service mais plutôt de généraliser le libre-service quand c'est nécessaire au travers de règles de contrôle et de gouvernance adaptés. Pour vous assurer de fournir le type de libre-service adéquat, votre solution de data lake doit disposer des fonctionnalités pour :

- **automatiser et éliminer le déploiement et la configuration complexes** pour que les utilisateurs accèdent rapidement aux ressources et aux environnements adéquats pour travailler avec les données ;
- **rendre les données accessibles** en déployant des outils intuitifs pour les utilisateurs métiers qui se servent des données du data lake pour prendre des décisions ;
- **mettre en place un libre-service gouverné** permettant un accès aux informations de l'entreprise en toute sécurité.
- **déployer** un processus décisionnel orienté données et permettre au plus grand nombre d'y prendre part.

LE CLOUD DATA WAREHOUSE : VERS L'INFINI ET AU DELÀ

Un cloud data warehouse de nouvelle génération offre une nouvelle façon de penser l'évolution du stockage et du traitement des données. Vous disposez désormais de ressources quasi illimitées que vous utilisez à votre guise : vous payez uniquement pour la capacité de traitement utilisée, au mois, au jour ou à l'heure. Vous apprécierez la même flexibilité avec le stockage, qui pourra être ajusté presque instantanément en fonction des demandes utilisateurs.

Les capacités d'un data lake vous permettent de gérer n'importe quel volume de données, charge de travail et imprévu. Vous pouvez garantir des prestations plus rapides à un plus grand nombre d'utilisateurs et fournir différentes vues de données en fonction de besoins spécifiques – tout cela grâce à une seule plateforme. Il existe plusieurs moyens de s'adapter facilement et de manière dynamique, sans temps de latence, aux pics et aux creux de la demande, grâce au redimensionnement et à l'utilisation concurrente (approche scale-up, scale-down et scale-out) :

- **redimensionner le stockage** (scale-out) pour conserver de grands ensembles de données disparates à moindre coût et indépendamment des cycles de traitement ;
- **redimensionner (scale-up et scale-down) en temps réel les ressources de traitement** pour faire face à des augmentations rapides des activités, par exemple dans le cas de requêtes des utilisateurs sur des volumes de données plus importants que d'habitude ;
- **redimensionner facilement un cluster existant** si vous souhaitez garder un contrôle strict sur les ressources et les coûts de traitement ;
- **garder un cluster de traitement prédéfini en mode suspendu**, mais prêt à redémarrer dans le cas d'un événement régulier nécessitant un redoublement des ressources informatiques.
- **lancer automatiquement un cluster supplémentaire pour gérer des utilisateurs simultanés**, sans sacrifier les performances lorsque leur nombre augmente. Quand la charge de travail diminue et les requêtes sont remises à niveau, le deuxième cluster ralentira automatiquement.

Avec toutes ces options, vous ne paierez que pour les ressources que vous utilisez, et le casse-tête d'ajouter des ressources supplémentaires à la volée, typique des technologies traditionnelles, ne sera plus qu'un lointain souvenir.

SÉCURITÉ INTÉGRÉE POUR VOTRE DATA LAKE

La sécurité est toujours d'une importance primordiale, surtout lorsque des informations à caractère personnel sont stockées dans le data lake. Correctement gérées, les mesures de sécurité fournies par un cloud data warehouse peuvent se révéler beaucoup plus efficaces et abordables que celles assurées par votre infrastructure de sécurité.

Quelles sont les caractéristiques de ce type de cloud data warehouse prenant en charge la sécurité ?

- Il intègre des protocoles de sécurité dans le respect des normes ;
- Il assure un contrôle des accès granulaire aux opérations et aux données ;
- Il fournit cette technologie à des milliers de clients, par défaut ;
- Il est constamment à jour avec les meilleures pratiques de sécurité, y compris l'utilisation du cryptage AES-256 et la conformité avec les normes SOC 2 Type 2 ;
- Il est certifié pour répondre aux exigences spécifiques de certains secteurs, telles que la législation sanitaire (HIPAA) et la norme de sécurité de l'industrie des cartes de paiement (PCI DSS).





Inversez la tendance avec l'analyse prédictive

PILOTEZ LE CHANGEMENT A L'AIDE DE NOUVEAUX INDICATEURS DE PERFORMANCE

Les entreprises sont en recherche constante de moteurs de croissance.

Au lieu d'utiliser les indicateurs historiques, elles veulent maintenant s'appuyer sur les indicateurs utiles à la direction de l'entreprise. Le data lake apporte une dimension prédictive et prescriptive à vos données. Qu'il s'agisse de sources de données internes ou externes, essentielles ou marginales par rapport aux activités de votre entreprise, le potentiel de ces données pourra être révélé à des fins d'analyses.

Un data lake construit sur un cloud data warehouse, permet à l'entreprise d'acquérir rapidement de nouvelles perspectives capables d'entraîner un changement profond.

À partir de là, les utilisateurs métiers et les data scientists peuvent se plonger dans une exploration plus approfondie, sans avoir à se connecter à de nouvelles sources de données ou exécuter des transferts de données complexes. Ils bénéficient d'un accès à des jeux de données jusqu'alors inexploités.

GÉREZ LA PRÉVISIBILITÉ

L'analyse prédictive est aujourd'hui la préoccupation majeure des organisations centrées sur la donnée. Un data lake moderne simplifie l'accès à une énorme quantité de données utilisées pour libérer toute la puissance prédictive et prescriptive d'une organisation et répondre ainsi à des questions telles que :

- Quand un équipement critique tombera-t-il en panne ?
- Quels emplacements publicitaires généreront plus de ROI chaque jour ?
- Comment optimiser les itinéraires des camions en fonction d'un changement météo annoncé ?
- Combien de paquets de cacahuètes devraient être stockés sur un vol donné ?
- Et des millions d'autres encore.

Grâce au data lake de nouvelle génération, l'analyse prédictive permet de découvrir d'immenses opportunités dans les données.



Pour approfondir

UN DATA LAKE NAVIGABLE, GOUVERNÉ ET RAPIDE

Vous pouvez dès aujourd'hui atteindre le plein potentiel pour votre data lake, tout en maintenant un contrôle et une gouvernance importants. La plateforme de data lake la plus adaptée est le data warehouse moderne dans le Cloud intégré à une solution de gestion de données de qualité. Votre entreprise pourra ainsi tirer parti d'un potentiel illimité de nouvelles connaissances.

Pour plus de renseignements sur la façon de construire le data lake optimal dès aujourd'hui, visitez snowflake.net et talend.com.

À propos de Snowflake

Snowflake est le seul data warehouse conçu pour le Cloud. Snowflake apporte la performance, la simplicité d'utilisation et l'utilisation concurrente pour stocker et analyser toutes les données existantes d'une organisation dans un même endroit. La technologie Snowflake combine la puissance d'un data warehouse, la flexibilité des plateformes Big Data, l'élasticité du Cloud et le partage de données en temps réel à un coût très inférieur à celui des solutions traditionnelles. Snowflake : vos données, sans limites. En savoir plus sur snowflake.net.

À propos de Talend

Talend est un leader de nouvelle génération dans les solutions d'intégration Big Data et Cloud, qui aide les entreprises à adopter une approche orientée données en rendant ces dernières plus accessibles, en améliorant leur qualité et en les plaçant là où elles sont nécessaires pour permettre la prise de décisions en temps réel. Grâce à Data Fabric, la plateforme d'intégration open source, native et unifiée de Talend, les clients accueillent les dernières innovations et les adaptent en fonction de l'évolution de leurs exigences.