



# 11 meilleures pratiques pour migrer vers un lac de données dans le cloud



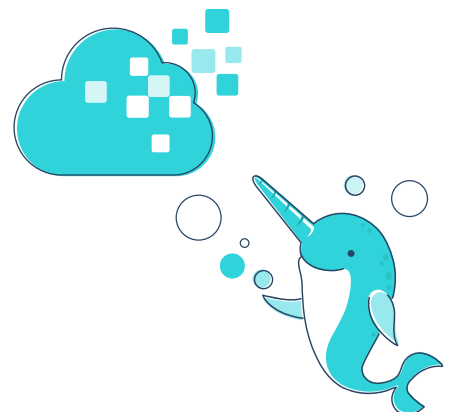
# Introduction

Les lacs de données sont devenus le lieu d'arrivée principal des données professionnelles. Cela est dû au fait qu'ils donnent aux entreprises un référentiel unique et consolidé pour toutes les données et tous les formats de données, de sorte qu'il est possible d'exploiter, de comprendre et d'intervenir efficacement sur les informations commerciales puissantes.

Toutefois, dès aujourd'hui, de nombreuses organisations, qui ont initialement déployé des lacs de données locaux, comprennent qu'un lac de données basé dans le cloud peut leur apporter un mélange irrésistible d'atouts, tels qu'un coût total de possession (TCO) réduit, des performances supérieures, une facilité d'utilisation et de gestion accrue, auquel ils n'ont pas accès avec leur approche sur site.

La migration d'un lac de données sur site vers un lac de données dans le cloud est très favorable. Toutefois, la migration s'accompagne d'un certain nombre de soucis.

**Nous pouvons vous aider. Voici 11 meilleures pratiques à suivre pour bénéficier d'une migration des données dans le cloud mieux réussie, rationalisée et transparente.**



# 11 meilleures pratiques pour migrer vers un lac de données dans le cloud

## PRÉ-MIGRATION

#1

### SÉLECTIONNEZ LE MEILLEUR FOURNISSEUR DE CLOUD EN FONCTION DE VOS BESOINS SPÉCIFIQUES EN TERMES DE LAC DE DONNÉES.

Pour évaluer, puis sélectionner le fournisseur de cloud qui répond le mieux aux besoins spécifiques de votre entreprise en matière de lac de données, veillez à inclure les critères suivants dans votre évaluation :

- + **Résultats à obtenir pour chaque cas d'utilisation.** Recherchez-vous un moteur de requête haute performance pour l'exploration de données, l'ingénierie de fonctionnalités ou l'ingénierie de données ? Autrement dit, quels sont les résultats exacts que vous souhaitez obtenir pour chaque cas d'utilisation spécifique que vous devez gérer ?
- + **Capacités.** Quel fournisseur de cloud offre la ou les spécialités particulières dont vous avez besoin, c'est-à-dire quel moteur prend le mieux en charge les charges de travail que vous avez besoin d'exécuter, telles que les analyses plus rapides ?
- + **Interopérabilité.** De quel niveau d'interopérabilité avez-vous besoin par rapport à celui qui est proposé par le fournisseur ? Même si le niveau exact d'interopérabilité requis sera très spécifique à votre cas d'utilisation, veillez à comparer le niveau d'interopérabilité proposé par chaque fournisseur de cloud à votre environnement actuel pour comprendre :
  - comment le stockage dans un lac de données dans le cloud du fournisseur communiquera avec les autres composants de votre lac de données dans le cloud ;
  - si, au minimum, le fournisseur peut prendre en charge le même niveau d'interconnectivité qui existe dans votre lac de données sur site ;
  - s'il existe des applications/services de lac de données qui communiquent entre eux ;
  - si vos applications continueront à bénéficier du même niveau d'interopérabilité une fois qu'elles seront déployées dans une nouvelle architecture de cloud ; et
  - s'il est possible et facile de connecter au lac de données dans le cloud du fournisseur tout autre service (dans le cloud comme hors du cloud) que vous pourriez déployer à l'avenir.
- + **Migration des données.** Sera-t-elle facile et combien de temps faudra-t-il pour migrer vos données vers le stockage du lac de données dans le cloud du fournisseur ? Comment devrez-vous gérer les données une fois qu'elles auront migré ?
- + **Stockage.** Le fournisseur de cloud répond-il aux exigences de votre cas d'utilisation en termes de disponibilité et de durabilité du stockage ? Considérez les abstractions de niveau objet-à-fichier, les différents degrés de redondance que votre sécurité peut nécessiter et si le fournisseur propose une disponibilité élevée et régulière.

- **Prix.** Comment utiliser la flexibilité, l'évolutivité et le modèle de tarification à l'utilisation du fournisseur de cloud à votre avantage ? Tenez compte des éléments suivants :
  - est-il plus avantageux de déployer immédiatement toutes les ressources de cloud du fournisseur, ou plutôt de commencer petit et d'augmenter progressivement au fil des besoins ?
  - le niveau de tolérance aux défaillances de votre budget et la demande de besoins, au-delà des niveaux minimum qu'offrent même les meilleurs tarifs (c'est-à-dire qu'un lac de données dans le cloud qui doit être répliqué sur plusieurs sites à travers le monde sera plus coûteux) ; et
  - l'impact des éventuels autres facteurs, tels que la vitesse d'accès dont vous avez besoin, par exemple, sur les frais que vous aurez à payer.

## Comparaison entre les services de cloud

Service de cloud	Ingestion	Stockage	Traitement	Analyses
<b>Microsoft Azure</b>	Azure Data Factory Azure Stream Analytics Apache Sqoop Azure Powershell Azure Portal Adlcopy DistCp	ADLS Blob Storage ADLS Gen2	HDInsight Azure SQL Data Warehouse HDinsight Storm	Data Lake Analytics
<b>Amazon Web Services</b>	Amazon Kinesis Amazon Snowball Amazon Storage Gateway	Amazon S3	AWS Glue Amazon Glacier	Amazon Athena Amazon EMR Amazon Redshift
<b>Google Cloud Platform</b>	Cloud Pub/Sub Cloud Dataflow Cloud Storage Transfer Service	Google Cloud Storage	Cloud Datalab Cloud Dataprep	Big Query Cloud Data Proc Cloud Bigtable

## #2

### ANALYSEZ ET MINIMISEZ L'IMPACT SUR LES UTILISATEURS.

Pour comprendre pleinement et minimiser l'impact négatif de la migration de votre lac de données vers le cloud sur les utilisateurs :

- déterminez quels applications, systèmes, départements et employés utilisent le lac de données sur site et la manière dont ils l'utilisent ;
- déterminez les éventuels changements qui se produiront lorsque le lac de données sera transféré dans le cloud, afin de pouvoir estimer tous les coûts implicites/explicites qui seront encourus et développez (ou mettez à jour) la documentation et les supports de formation si nécessaire ;
- déterminez le moment optimal pour faire migrer les cas d'utilisation et/ou les charges de travail spécifiques, en évitant les heures d'utilisation intensive ou critique ; et
- estimez la portée et la durée des interruptions potentielles de service et élaborer un plan pour les communiquer de manière proactive aux utilisateurs à l'avance.

## #3

### DÉVELOPPER ET METTRE EN ŒUVRE UNE STRATÉGIE DE GOUVERNANCE ET DE SÉCURITÉ DES DONNÉES.

Pour protéger les données et garantir une conformité continue aux obligations gouvernementales, développez et mettez en œuvre une stratégie de gouvernance et de sécurité des données qui :

- + respecte les réglementations et les lois du pays dans lesquelles les données de votre cloud seront stockées, ainsi que celles du pays d'origine de votre entreprise ;
- + protège les données de votre entreprise contre les préjudices ou pertes inattendus ;
- + sauvegarde les données avant la migration et prévoit la manière dont les données seront restaurées si nécessaire ;
- + identifie les rôles qui peuvent accéder au lac de données dans le cloud et spécifie les autorisations précises à utiliser pour octroyer cet accès ; et
- + décrive une nouvelle architecture de sécurité qui ne soit ni trop simpliste (entraînant un risque d'usurpation de données), ni trop complexe (ce qui augmente le risque lorsque les utilisateurs tentent de contourner les mesures de sécurité pour effectuer leur travail).

#### Info rapide

Selon le « **contrat de responsabilité partagée** » établi par les fournisseurs de cloud, **votre entreprise doit seulement sécuriser les données qu'elle envoie dans le cloud**. Toutes les autres tâches de sécurité relèvent de la responsabilité du fournisseur de cloud.

## #4

### FAIRE ÉVOLUER LES PIPELINES DE DONNÉES ET AUTANT D'APPLICATIONS QUE POSSIBLE POUR QU'ILS SOIENT CLOUD-NATIFS.

Pour optimiser les avantages de la nouvelle architecture de lac de données et atteindre une efficacité optimale, utilisez cette migration dans le cloud comme une opportunité pour :

- + évaluer les autres pipelines de données et applications existants qui bénéficieraient du passage dans le cloud au cours de la migration ;
- + identifier une conception plus moderne et plus flexible pour toutes les applications ; et
- + passer en revue les autres investissements liés au lac de données par rapport à votre nouvelle architecture de cloud.

# #5

## CRÉER UN PLAN DE MIGRATION DÉTAILLÉ.

Pour contribuer à minimiser les problèmes que vous rencontrerez lors de la migration de votre lac de données dans le cloud, établissez et respectez un plan de migration détaillée contenant des listes de :

- + sous-tâches englobant tous les éléments discutés ci-dessus, ainsi que l'heure de leur application, les échéances et les mesures de succès ;
- + outils existants qui vous permettent de capturer, puis d'exporter les éventuelles sémantiques de données sur site, tels que les vues, les indicateurs clés de performance (KPI) et la logique commerciale, afin de pouvoir les appliquer au lac de données dans le cloud (en revisitant la cartographie des jeux de données physiques si nécessaire) ;
- + outils existants qui fournissent des modèles opérationnels et des expériences utilisateur cohérents pour les déploiements sur site et dans le cloud ;
- + procédures à exécuter après la migration ; et
- + toutes les parties prenantes qui interviendront à chaque étape de la migration.

### Suggestions de parties prenantes à impliquer

Pré-migration	Pendant la migration	Après la migration
<ul style="list-style-type: none"><li>• Directeur de l'informatique (CIO)</li><li>• Directeur des données (CDO)</li><li>• Directeur de la technologie (CTO)</li><li>• Le responsable de la décision finale</li></ul>	<ul style="list-style-type: none"><li>• Un chef de projet (qui communiquera les mises à jour, coordonnera les activités et suivra la progression)</li><li>• Des ingénieurs en données familiarisés avec les données (qui identifieront l'ordre de leur déplacement)</li><li>• Des spécialistes de l'informatique (SME) familiarisés avec l'infrastructure de données (qui superviseront et assisteront la migration)</li><li>• Des architectes/ingénieurs de données possédant une expérience de la migration dans le cloud (qui contribueront à identifier de manière proactive et à gérer les obstacles potentiels)</li><li>• Les consommateurs de données (qui corroboreront les résultats tout au long du processus)</li></ul>	<ul style="list-style-type: none"><li>• Les analystes de la veille économique (BI)</li><li>• Les ingénieurs de données</li><li>• Les responsables informatiques</li><li>• Les éventuelles autres fonctions possédant une compréhension approfondie du lac de données sur site, des processus de pipeline actuels, des infrastructures de système de source de données et des données elles-mêmes, et qui pourront aider à la validation des résultats finaux</li></ul>

## PENDANT LA MIGRATION

#6

### SAUVEGARDER LES DONNÉES.

Pour empêcher la perte de données pendant la migration, sauvegardez vos données (et leurs fichiers et paramètres de configuration) avant de commencer le processus de migration.

#7

### TRANSFÉRER LES DONNÉES PAR INCRÉMENTS.

Après avoir établi toutes les connexions nécessaires avec votre nouveau fournisseur de lac de données dans le cloud, vous êtes prêt à commencer le processus de migration.

Pour minimiser les interruptions et assurer une haute disponibilité des données critiques pour l'entreprise, commencez par une « étape pilote » qui ne transférera que quelques charges de travail et/ou sources de données par cas d'utilisation.

Ensuite, donnez la priorité à la migration des charges de travail non-critiques dont l'interruption ne perturbera pas les opérations quotidiennes importantes (c'est-à-dire qu'il ne faut pas faire migrer des charges de travail financières à la fin d'un trimestre).

#8

### TESTEZ SOIGNEUSEMENT LA MIGRATION AVANT DE TRANSFÉRER LES DONNÉES RESTANTES.

Avant de migrer toutes les données restantes, procédez à un exercice exhaustif d'assurance qualité pour valider l'intégrité des données, les connexions de données et les interactions adéquates entre tous les systèmes connectés au nouveau lac de données dans le cloud.

---

### N'oubliez pas de former le formateur

Veillez à laisser aux administrateurs chargés de gérer votre nouveau lac de données dans le cloud un temps largement suffisant pour se familiariser et s'adapter à ce nouveau cloud, afin qu'ils soient préparés à assister efficacement les utilisateurs du lac de données dans le cloud une fois que le déploiement commencera.

---



## APRÈS LA MIGRATION

#9

### TESTEZ À NOUVEAU SOIGNEUSEMENT LA MIGRATION.

Une fois toutes les données migrées dans le nouveau lac de données dans le cloud, et avant de le déployer pour un public plus large, optimisez les paramètres du lac de données dans le cloud, puis vérifiez l'accessibilité et l'intégrité des données.

Pour ce faire, testez soigneusement un important échantillon de données qui représente différentes plages horaires, en utilisant des consommateurs de données expérimentés et attentifs aux détails pour valider la correspondance entre les données d'historique et les résultats obtenus avec la nouvelle plateforme.

#10

### INFORMEZ ET ÉDUQUEZ TOUS LES UTILISATEURS ET TOUTES LES PARTIES PRENANTES AU SUJET DU NOUVEAU LAC DE DONNÉES DANS LE CLOUD.

Pour faciliter l'adoption appropriée et le succès de la nouvelle plateforme de lac de données dans le cloud :

- + préparez les utilisateurs à la migration à l'avance en leur apprenant à utiliser le nouveau lac de données dans le cloud ;
- + orientez l'essentiel de vos efforts de formation au lac de données dans le cloud vers le développement des compétences du personnel technique, mais informez et éduquez également à la fois les utilisateurs et les parties prenantes au sujet du projet ; et
- + partagez les mises à jour au sujet de la progression du projet au quotidien, afin que les parties prenantes et les consommateurs des données fassent confiance et s'adaptent plus rapidement à la nouvelle plateforme dans le cloud.

#11

### NE FERMEZ PAS IMMÉDIATEMENT VOTRE LAC DE DONNÉES SUR SITE.

Pour atténuer les interruptions potentielles, ne fermez pas trop tôt votre lac de données sur site tant que toutes les parties prenantes n'ont pas complètement approuvé le nouveau lac de données dans le cloud. Le temps exact qu'il vous faudra pour conserver les données héritées sur site disponibles dépendra de :

- + la taille de votre entreprise ;
- + la quantité de données à déplacer ; et
- + le temps qu'il faudra aux parties prenantes pour vérifier les données migrées.

Par exemple, les parties prenantes pour une petite à moyenne entreprise qui ne déplace que sa charge de travail financière vers le cloud peuvent n'avoir besoin que d'un ou deux mois pour vérifier le bon fonctionnement du nouveau pipeline. En revanche, les parties prenantes d'une entreprise de plus de 1 000 employés qui déplace l'intégralité de sa charge de travail vers un lac de données dans le cloud peuvent avoir besoin d'un an ou plus pour vérifier leurs résultats.



# Votre fiche de travail pour une migration de lac de données dans le cloud inspirée par les meilleures pratiques

## PRÉ-MIGRATION

### #1 Sélectionnez le meilleur fournisseur de cloud en fonction de vos besoins spécifiques en termes de lac de données.

- ☐ Identifiez les résultats dont vous avez besoin pour chaque cas d'utilisation.
- ☐ Déterminez l'expertise/les capacités spécialisées de chaque fournisseur de cloud.
- ☐ Comparez l'interopérabilité dont vous avez besoin à celle qui est proposée par chaque fournisseur de cloud.
- ☐ Déterminez le temps qu'il faudrait et la facilité de la migration des données vers le cloud de chaque fournisseur.
- ☐ Déterminez la gérabilité de vos données une fois qu'elles seront migrées dans le service de chaque fournisseur de cloud.
- ☐ Déterminez quels fournisseurs de cloud répondent aux exigences de vos cas d'utilisation en matière de disponibilité du stockage et de durabilité.
- ☐ Évaluez la structure des prix/tarifs de chaque fournisseur de cloud.

### #2 Analysez et minimisez l'impact sur les utilisateurs.

- ☐ Identifiez les utilisateurs/départements et les systèmes/applications qui utilisent des données existantes du lac de données sur site, et dans quel but.
- ☐ Calculez tous les coûts implicites et explicites qui seront encourus lors de la migration.
- ☐ Déterminez s'il faut ou non créer et/ou mettre à jour de nouvelles documentations et supports de formation pour les utilisateurs.
- ☐ Définissez l'ordre optimal de migration de cas d'utilisation et/ou de charges de travail spécifiques.
- ☐ Créez un plan pour communiquer de manière proactive aux utilisateurs l'ampleur et la date des perturbations potentielles du service.

### #3 Développez et mettez en œuvre une stratégie de gouvernance et de sécurité des données.

- ☐ Assurez-vous que votre stratégie respecte les règlements et les lois du pays d'origine de votre entreprise.
- ☐ Assurez-vous que votre stratégie respecte les réglementations et les lois du pays dans lequel les données de votre cloud seront stockées.
- ☐ Définissez la manière dont les données de l'entreprise seront protégées contre les préjudices ou les pertes imprévus pendant la migration.
- ☐ Déterminez comment vous sauvegarderez les données avant le début de la migration.
- ☐ Documentez la manière dont les données seront restaurées si certaines sont perdues ou endommagées pendant la migration.
- ☐ Identifiez les rôles autorisés à accéder au lac de données dans le cloud.
- ☐ Spécifiez les autorisations précises à utiliser pour octroyer différents types d'accès au lac de données dans le cloud.
- ☐ Définissez une architecture de sécurité d'une rigidité appropriée.

### #4 Faites évoluer les pipelines de données et autant d'applications que possible pour qu'ils soient cloud-natifs.

- ☐ Évaluez les autres pipelines de données et applications qui pourraient bénéficier du déplacement vers le cloud lors de la migration.
- ☐ Identifiez une conception plus moderne et plus flexible pour toutes les applications.
- ☐ Passez en revue les autres investissements liés au lac de données par rapport à votre nouvelle architecture de cloud.

### #5 Créez un plan de migration détaillé.

- ☐ Détaillez toutes les sous-tâches et les dates, échéances et mesures de succès correspondantes.
- ☐ Documentez les outils existants qui fournissent des modèles opérationnels et des expériences utilisateur cohérents pour les déploiements sur site et dans le cloud.
- ☐ Documentez les éventuelles sémantiques de données sur site qui peuvent être exportées depuis les outils existants (et si/comment leurs jeux de données physiques doivent être remappés).
- ☐ Répertoirez les procédures post-migration à exécuter, et dans quel ordre.
- ☐ Identifiez toutes les parties prenantes à impliquer à chaque phase de la migration.

## PENDANT LA MIGRATION

### #6 Sauvegardez les données.

- ☐ Sauvegardez toutes les données.
- ☐ Sauvegardez tous les fichiers et paramètres de configuration pour ces données.

### #7 Transférez les données par incréments.

- ☐ Commencez par transférer seulement quelques charges de travail et/ou sources de données par cas d'utilisation.
- ☐ Donnez la priorité à la migration des charges de travail non critiques.

### #8 Testez soigneusement la migration avant de transférer les données restantes.

- ☐ Validez l'intégrité des données initiales qui sont migrées.
- ☐ Vérifiez le bon fonctionnement de toutes les connexions de données.
- ☐ Confirmez les bonnes interactions entre tous les systèmes connectés au nouveau lac de données dans le cloud.

## APRÈS LA MIGRATION

### #9 Testez à nouveau soigneusement la migration.

- ☐ Optimisez les paramètres du lac de données dans le cloud.
- ☐ Vérifiez que les valeurs historiques correspondent aux résultats obtenus de la nouvelle plateforme.

### #10 Informez et éduquez tous les utilisateurs et toutes les parties prenantes.

- ☐ Informez et éduquez tous les utilisateurs et toutes les parties prenantes.
- ☐ Formez le personnel technique de façon plus complète.
- ☐ Partagez des mises à jour quotidiennes sur le projet avec tous les utilisateurs et toutes les parties prenantes.

### #11 Ne fermez pas immédiatement votre lac de données sur site.

- ☐ Gagnez l'approbation totale de toutes les parties prenantes sur le nouveau lac de données dans le cloud avant de fermer votre lac de données sur site.

# Conclusion

Il ne fait aucun doute que les lacs de données dans le cloud offrent une multitude d'avantages par rapport à leurs équivalents sur site. Pourtant, de nombreuses entreprises n'ont pas encore franchi le pas en raison de préoccupations obsolètes au sujet du maintien de la sécurité des données, de la préservation de la continuité de l'activité et du simple achèvement de la migration elle-même.

Ensuite, en respectant les meilleures pratiques présentées dans cet e-book, vous serez en mesure de faire migrer avec succès et en toute transparence les données de votre entreprise depuis un lac de données sur site vers le cloud, et de bénéficier plus rapidement de tous les avantages qui en résultent.



---

Découvrez comment seul Dremio peut offrir à la fois une cohérence incroyable des requêtes, une efficacité de l'infrastructure du cloud et une accessibilité simplifiée, pour toutes les données et en permanence, moyennant un coût moindre.

Lire le [Guide de l'architecture Dremio](#).

---



## À PROPOS DE DREMIO

Dremio propose des requêtes rapides comme l'éclair et une couche de sémantique en libre-service directement sur votre stockage de lac de données. Aucun déplacement de données vers des entrepôts de données exclusifs, pas de cubes, pas de tables d'agrégation ni d'extraits. Simplement la flexibilité et le contrôle pour les architectes de données et le libre-service pour les consommateurs de données.

## Déployer Dremio

[En savoir plus sur dremio.com](#)

CONTACTER LE SERVICE COMMERCIAL  
[contact@dremio.com](mailto:contact@dremio.com)