



5 QUESTIONS À SE POSER AVANT DE MIGRER VERS SNOWFLAKE

Guide d'évaluation et de préparation à la migration



CHAMPION
GUIDES

eBOOK

TABLE DES MATIÈRES

- 3** Introduction
- 4** Mon entrepôt offre-t-il des performances d'analyse rapides ?
- 5** Ai-je des difficultés dans la gestion de mes charges de travail ?
- 6** Ma plateforme est-elle réellement évolutive ?
- 7** Ai-je la possibilité de garantir un partage de données rapide ?
- 8** Est-ce que je paie pour plus de ressources que ce dont j'ai réellement besoin ?
- 9** Pour aller plus loin
- 10** À propos de Snowflake

INTRODUCTION

Il existe de nombreuses raisons susceptibles d'inciter les utilisateurs de bases de données traditionnelles à envisager la migration de leur plateforme existante vers une alternative cloud comme Snowflake. Comme n'importe quel logiciel, les bases de données traditionnelles présentent des limites. Lenteur et inefficacité des requêtes, problèmes d'évolutivité et explosion des coûts de licence en cas de croissance de l'activité sont autant d'inconvénients évoqués par de nombreux utilisateurs. Ceux qui ont eu à composer avec ces systèmes complexes, leurs difficultés de maintenance et leurs problèmes de collaboration trouveront dans le Data Cloud Snowflake une alternative des plus convaincantes.

En effet, le Data Cloud Snowflake se distingue d'autres services grâce à plusieurs fonctionnalités clés :

- **Une plateforme unique** : prise en charge d'un ensemble complet de cas d'usage pour tout type d'organisation, à partir d'une même copie de données et dans une variété de formats de données (structurées, semi-structurées et non structurées), de langages (SQL, Python, Java et Scala) et de modèles d'architecture (entrepôt de données, data lake, solution hybride de ces deux modèles ou encore data mesh).
- **Tout fonctionne, tout simplement** : l'architecture différenciée et le modèle de prestation de service de Snowflake garantissent que la plateforme fonctionne, tout simplement. Grâce à différentes stratégies d'automatisation et d'optimisation, nos clients bénéficient en permanence de données efficaces, gouvernées et sécurisées à grande échelle. Snowflake fournit également des performances et une simultanéité de pointe dès sa mise en service, ce qui libère les utilisateurs de toute préoccupation de conflit de ressources ou de réglage. Enfin, grâce à Snowgrid, la couche technologique cross-cloud unique de Snowflake, les écosystèmes d'entreprise sont connectés entre les régions et les clouds à l'échelle mondiale, ce qui renforce la collaboration cross-cloud, la gouvernance des données et la continuité de l'activité.

- **Une solution rentable avec des ressources dimensionnées au plus juste** : Snowflake améliore continuellement son rapport prix/performance. La capacité de calcul élastique quasi instantanée propriétaire de Snowflake élimine tout surdimensionnement inutile ainsi que tout retard coûteux lors des opérations de démarrage et de mise à l'échelle des clusters. Grâce à des mises à jour transparentes apportées en continu à son moteur de performance, le rapport prix/performance de cette solution ne cesse de s'améliorer au fil du temps. De plus, avec la tarification à la consommation en temps réel de Snowflake, les utilisateurs ont la garantie de payer uniquement pour ce qu'ils utilisent, et sont assurés que toute amélioration de performances se traduit immédiatement par une réduction de coûts.

Fort de milliers de clients, de dizaines de prestataires partenaires et de plus d'une décennie d'expérience, Snowflake a acquis une compréhension fine du processus de migration et des facteurs qui en assurent la réussite. Nous en sommes venus à identifier pourquoi les entreprises souhaitent opérer un changement et quelles sont les questions les plus importantes à considérer avant de se lancer dans la migration de leur base de données actuelle vers Snowflake. Dans cet eBook, nous détaillons cinq questions clés que les entreprises doivent se poser pour évaluer leur besoin et leur état de préparation vis-à-vis de leur projet de migration.

- 1 **Mon entrepôt offre-t-il des performances d'analyse rapides ?**
- 2 **Ai-je des difficultés dans la gestion de mes charges de travail ?**
- 3 **Ma plateforme est-elle réellement évolutive ?**
- 4 **Ai-je la possibilité de garantir un partage de données rapide ?**
- 5 **Est-ce que je paie pour plus de ressources que ce dont j'ai réellement besoin ?**

1 : MON ENTREPÔT OFFRE-T-IL DES PERFORMANCES D'ANALYSE RAPIDES ?

Des performances d'analyse rapides sont essentielles pour assurer des analyses de données efficaces, des prises de décisions éclairées reposant sur ces données et la compétitivité des entreprises. Ces performances améliorées permettent aux analystes de travailler efficacement, d'adapter le volume de données gérées et d'identifier rapidement les erreurs.

Les technologies d'analyse ont considérablement évolué au cours de la dernière décennie, ce qui a permis d'accélérer de manière significative les délais de rentabilisation. Grâce aux analyses basées sur le cloud, les organisations peuvent désormais disposer d'une puissance de calcul sur de courtes périodes, ce qui facilite l'exécution des projets et permet même aux petites entreprises de tirer parti des avantages d'un entrepôt de données. Face à la prolifération des options de bases de données, la vitesse est devenue un facteur de différenciation majeur.

Les bases de données traditionnelles sont souvent des systèmes transactionnels par nature, avec lesquels les ingénieurs peinent à fournir des charges de travail analytiques et la vitesse nécessaire au traitement de gros volumes de données.

C'est pourquoi les bases de données de type OLTP (Online Transaction Protocol) sont souvent optimisées pour fonctionner avec un ensemble spécifique de lignes. Cette méthode permet d'écrire facilement une requête pour extraire les données nécessaires. Néanmoins, les requêtes portant sur des millions de lignes demandent beaucoup de temps, étant donné que chaque ligne doit être lue sur un disque et traitée par l'unité centrale (CPU) avant la lecture de la ligne suivante. Ce manque d'efficacité, associé au manque d'expérience de certains développeurs avec des langages comme SQL pour rédiger des requêtes, peut entraîner des problèmes majeurs en cas de tentative de redimensionnement.

En revanche, lorsque les données sont chargées avec Snowflake, elles sont réorganisées dans un format en colonnes optimisé et compressé. La plateforme Snowflake gère le stockage, de l'organisation et la taille des fichiers à la compression et aux métadonnées. Comme les bases de données en colonnes utilisent moins de mémoire pour renvoyer des données, il est possible d'en stocker davantage, ce qui réduit les délais de traitement des requêtes. Cette structure est non seulement primordiale pour exploiter des ensembles de mégadonnées (Big Data), mais elle est également indispensable à toute analyse de données accélérée.



2 : AI-JE DES DIFFICULTÉS DANS LA GESTION DE MES CHARGES DE TRAVAIL ?

Lorsque des requêtes complexes de longue durée sont exécutées sur une même machine on-premise, les capacités de calcul et de stockage sont étroitement associées sans être séparées. Dans cette structure, le stockage peut être prioritaire aux dépens du calcul, ou vice-versa. Ce partage de ressources peut générer des problèmes d'évolutivité et de rapidité.

De fait, les data engineers se voient contraints d'exécuter leurs charges de travail à certaines heures de la journée, car ils ne peuvent pas toutes les traiter simultanément à cause de ressources de calcul limitées. Une simple erreur de planification peut détériorer le serveur, ou même le rendre inutilisable. Il s'agit là d'une pression considérable.

Ces dernières années, l'analyse des données a délaissé le simple traitement des transactions et de plus en plus de plateformes, dont Snowflake, dissocient les fonctions de calcul et de stockage. Les avantages sont triples :

- **Évolutivité.** L'ancien modèle, dans lequel les capacités de calcul et de stockage sont étroitement liées, complique le dimensionnement des bases de données. En effet, lorsque des nœuds de base de données sont ajoutés, les données dans le cluster doivent être « rééquilibrées ». Or, ce processus est gourmand en ressources et nécessite souvent un temps d'arrêt. À l'inverse, lorsque le stockage et le calcul sont séparés, les deux peuvent évoluer indépendamment pour répondre aux différents cas d'usage, généralement en quelques secondes seulement.
- **Disponibilité.** Lorsque toutes les données sont stockées à distance, l'échec d'un nœud de calcul n'entraîne pas de conséquences graves. La récupération se fait rapidement et en toute simplicité, car les nœuds restants de n'importe quel cluster peuvent toujours accéder à toutes les données sur le réseau.
- **Coût.** La séparation des ressources de calcul et de stockage permet aux entreprises de ne payer que pour ce qu'elles consomment. Selon certaines estimations, cette tarification peut offrir **jusqu'à 66 % d'économies** sur les coûts de calcul dans des secteurs tels que le retail, qui connaît des périodes de pic de demande (p. ex. lors du Black Friday), sans pour autant avoir besoin d'un tel dimensionnement toute l'année. Cette structure élimine quasiment toute la complexité et tous les coûts liés aux copies de données, tout en fournissant une source de vérité unique.

Ce nouveau paradigme des plateformes data vous permet de faire appel à des ressources de calcul adaptées à n'importe quelle échelle pour interroger le stockage partagé, puisque les requêtes n'affectent pas les autres charges de travail, qui disposent de leurs propres ressources de calcul dédiées. Tout est une question d'allocation intelligente des ressources.



3 : MA PLATEFORME EST-ELLE RÉELLEMENT ÉVOLUTIVE ?

Toutes les organisations veulent se développer en s'appuyant sur des systèmes capables de suivre le rythme de leur croissance. Si les bases de données traditionnelles sont capables d'évoluer, cela nécessite généralement de supprimer le serveur existant pour en mettre en place un nouveau en l'espace d'une heure environ, ce qui met en même temps les activités à l'arrêt. Or, ces perturbations entraînent des temps d'arrêt côté client, c'est-à-dire une perte de revenus et des frustrations.

Dans le cas d'une mise à l'échelle horizontale, dès lors que l'on cherche à augmenter le nombre de requêtes simultanées, il devient souvent nécessaire de procéder à des mises à niveau matérielles onéreuses, dont le coût ne peut être adapté à la baisse. D'un autre côté, si une évolution verticale permet aux entreprises d'accroître la puissance de traitement d'un serveur ou d'un cluster unique, il existe un risque de perte de données en cas de défaillance matérielle ou logicielle. Par conséquent, les deux méthodes présentent des risques et des limitations quant à la mise à niveau des machines.

Le moteur de performances élastique de Snowflake a été conçu pour offrir une évolutivité fluide. Ainsi, quelques secondes suffisent le plus souvent pour permettre aux entreprises d'obtenir une simultanéité dès la mise en service, sans avoir à se soucier des conflits de ressources, des réglages ou de la gestion du système. De cette façon, les équipes data peuvent accomplir leurs tâches sans avoir à gérer les systèmes eux-mêmes. Snowflake peut prendre en charge un nombre quasi illimité d'utilisateurs, de tâches ou de données en assurant une isolation fiable des ressources sur plusieurs clusters.

Snowflake permet de redimensionner les entrepôts afin d'améliorer les performances des requêtes et de réduire les files d'attente, et ce, pratiquement n'importe quand, même pendant l'exécution et le traitement de requêtes. Il est possible de démarrer ou d'arrêter des entrepôts à tout moment, mais également d'en créer rapidement de nouveaux pour soutenir n'importe quel type d'activité. Cette flexibilité garantit une valeur économique maximale et permet aux utilisateurs d'optimiser en permanence le rapport prix/performance.

Snowflake intègre dans son infrastructure des fonctions clés de gestion, de réglage et de nettoyage pour que vous n'ayez plus à surveiller et gérer le système et pour réduire votre coût total de possession. Cette maintenance quasi nulle est une opération en ligne automatisée qui nécessite rarement de programmer des temps d'arrêt.



4 : AI-JE LA POSSIBILITÉ DE GARANTIR UN PARTAGE DE DONNÉES RAPIDE ?

Le partage de données garantit aux équipes métiers un accès plus large et en temps quasi réel aux données les plus pertinentes dont elles ont besoin pour accélérer la croissance de leur activité. Dans les anciens systèmes, il est très courant de voir des bases de données intégrer une logique supplémentaire dans leur pipeline de données pour permettre aux utilisateurs de copier des ensembles de données d'un environnement à un autre.

De nombreux utilisateurs de bases de données traditionnelles rencontrent des difficultés lorsqu'ils essaient de partager des ensembles de données importants par le biais de processus de partage traditionnels comme le protocole FTP (File Transfer Protocol), le protocole réseau standard utilisé pour le transfert de fichiers informatiques entre un client et un serveur sur un réseau d'ordinateurs. Le protocole FTP peut être complexe à utiliser via n'importe quel type de pare-feu et de nombreuses passerelles ont du mal à prendre en charge les connexions entrantes, ce qui engendre complications et retards.

Gérer des systèmes FTP traditionnels peut présenter des risques, car nombre d'entre eux ne permettent pas de chiffrer des données. En raison de ce manque de sécurité, l'expéditeur perd le contrôle des données qu'il envoie par FTP et ne peut guère décider qui accède aux données et pendant combien de temps les fichiers peuvent être utilisés. En outre, ces systèmes ont souvent peu à offrir en termes d'automatisation, obligeant les équipes à consacrer de nombreuses heures à gérer manuellement les transferts. Ces complexités s'ajoutent aux incohérences dans les bases de données, aux problèmes d'authenticité et de partage d'importants volumes de données, créant alors des processus d'échange de données frustrants, coûteux et chronophages.

C'est pourquoi Snowflake intègre désormais le Data Sharing aux autres fonctionnalités déjà prises en charge. Ainsi, les utilisateurs peuvent facilement extraire des ensembles de données d'autres comptes Snowflake et sélectionner des sous-ensembles de données à partager avec d'autres. Les utilisateurs n'ont donc plus besoin de créer des tâches ETL pour transférer des ensembles de données d'un entrepôt à un autre. Les autorisations liées à chaque ensemble de données sont simples à gérer et permettent aux seuls utilisateurs qui en ont besoin d'y accéder au moment où ils en ont besoin.

Snowflake a également la capacité de partager des téraoctets de données sans avoir à les déplacer physiquement, une solution qui vient s'ajouter à une série de caractéristiques conçues pour être fonctionnelles, tout simplement. Avec ces outils à leur disposition, les data engineers ne sont plus contraints de créer des pipelines entrants et sortants pour partager des données.

Conçue pour les entreprises data-driven modernes, la plateforme Snowflake a été développée en gardant toujours à l'esprit que les entreprises prospères doivent échanger des données pour innover.

- Partagez des données facilement et en toute sécurité entre les clouds, les entreprises, les équipes, les services, les fonctions et les unités commerciales. Si deux entreprises appartiennent à la même région de cloud Snowflake, ce partage de données devient presque instantané.
- Configurez facilement la sécurité et la gouvernance avec des autorisations et des rôles intégrés.
- Partagez des données, des vues et des tableaux de bord en fonction de vos besoins, afin d'améliorer la collaboration et la prise de décisions.
- Améliorez la planification et appuyez-vous sur des informations pour répondre aux besoins de vos clients, gérer vos ressources et réduire vos délais de commercialisation de nouveaux services.
- Répliquez des données entre différents fournisseurs de cloud à l'aide des fonctionnalités cross-cloud de Snowflake, qui permettent un partage de données fluide entre les clouds publics sans travail supplémentaire.

5 : EST-CE QUE JE PAIE POUR PLUS DE RESSOURCES QUE CE DONT J'AI RÉELLEMENT BESOIN ?

Dans la plupart des bases de données traditionnelles, les utilisateurs peuvent arrêter ou mettre en pause des serveurs lorsqu'ils ne les utilisent pas. Toutefois, il n'existe pas de solution satisfaisante pour surveiller et détecter des demandes d'utilisateurs en matière de serveurs uniques. Par conséquent, bon nombre d'entreprises les maintiennent en fonctionnement en permanence et en paient le prix. Si elles mettent un serveur en pause, elles sont facturées pour une heure complète, même s'il ne fonctionne que quelques minutes.

De plus, la pile technologique nécessaire pour prendre en charge un entrepôt de données moderne représente un investissement majeur. Les tarifs des technologies varient considérablement en fonction des besoins et des objectifs d'une organisation, mais les coûts ont tendance à augmenter au fil du temps et de la croissance. L'achat d'une quantité définie de stockage et de temps et de puissance de calcul entraîne des coûts annuels significatifs et nécessite beaucoup de gestion. Une équipe d'analyse peut donc y consacrer un temps précieux, qui serait mieux employé sur des tâches d'analyse de données pour en tirer un avantage concurrentiel.

Avec des ressources efficaces et rentables, les entreprises peuvent se concentrer sur leurs résultats, sans sacrifier les outils nécessaires à leur innovation. C'est pourquoi le modèle d'utilisation de Snowflake permet aux entreprises de ne payer que pour les ressources dont elles ont besoin, avec une structure axée sur la consommation plutôt que sur les pics.

Les ressources de calcul de Snowflake sont facturées à la seconde, avec un minimum d'une minute, ce qui réduit considérablement les coûts pour les entreprises. La vitesse offerte par Snowflake permet elle aussi des économies grâce à l'exécution instantanée des charges de travail, qui rend la tarification à la seconde particulièrement intéressante. Le prix de cette efficacité commence en moyenne à 23 USD par **téraoctet compressé automatiquement par mois**, et un stockage haut de gamme coûteux n'est plus nécessaire pour assurer une gestion efficace des données.

POUR ALLER PLUS LOIN

Alors que les organisations créent et consomment des données à une vitesse inédite, une gestion de données efficace ne doit pas ajouter de la complexité, mais au contraire apporter une simplification. Cette philosophie est au cœur de la conception de Snowflake. Cette valeur fondamentale de simplicité s'applique également au processus de migration.

Pour rationaliser vos bases de données et rendre le partage, l'évolutivité et l'analyse aussi fluides que possible, contactez l'équipe des [services professionnels Snowflake](#), qui pourra vous fournir des outils comme SnowConvert. Ce système a été conçu pour faciliter et accélérer la migration de tables, de vues, de procédures, de macros, d'index joints et de scripts propriétaires à partir de systèmes anciens vers Snowflake.

Prêts à changer ? Contactez l'équipe des services professionnels Snowflake pour découvrir comment Snowflake peut améliorer vos opérations.

Ressources supplémentaires

Migrer vers le Data Cloud en toute simplicité.

[En savoir plus sur l'initiative Migration accélérée de Snowflake.](#)





À PROPOS DE SNOWFLAKE

Snowflake permet à chaque organisation de mobiliser ses données grâce au Data Cloud Snowflake. Ses clients utilisent le Data Cloud pour réunir au même endroit leurs données silotées, analyser et partager en toute sécurité les données, propulser des applications de données et exécuter diverses charges de travail analytiques et d'IA/ML. Quel que soit l'endroit où se trouvent les données ou les utilisateurs, Snowflake offre une expérience unique qui s'étend sur plusieurs clouds et régions. Au 31 janvier 2024, des milliers de clients de nombreux secteurs, dont 691 des Forbes Global 2000 (G2K) de 2023, utilisent le Data Cloud Snowflake pour dynamiser leur activité.

En savoir plus sur snowflake.com/fr.



© 2023 Snowflake Inc. Tous droits réservés. Snowflake, le logo Snowflake et tous les autres noms de produits, de fonctionnalités et de services Snowflake mentionnés dans le présent document sont des marques déposées ou des marques commerciales de Snowflake Inc. aux États-Unis et dans d'autres pays. Tous les autres noms de marque ou logos mentionnés ou utilisés dans le présent document le sont uniquement à des fins d'identification et peuvent être des marques de commerce de leur(s) détenteur(s) respectif(s). Snowflake peut ne pas être associé à, ou être sponsorisé ou approuvé par, un tel détenteur.