

**GUIDE COMPARATIF**

# Les meilleurs Data Warehouses dans le Cloud d'entreprise

**Amazon, Azure, Google ou Snowflake**

**Qlik**  **LEAD WITH DATA™**



# L'ascension du Data Warehouse dans le Cloud

Les data warehouses font partie de l'analytics et du reporting d'entreprise depuis des décennies. Cependant, ils n'ont pas été conçus pour gérer la croissance spectaculaire que connaît la data aujourd'hui ou pour suivre le rythme effréné de l'évolution des besoins des utilisateurs finaux. Avec l'émergence des data warehouses dans le cloud, la révolution est en marche.

Les data warehouses dans le cloud offrent aux entreprises de toutes tailles des avantages et une flexibilité jusqu'alors impossibles. Libérées des contraintes des datacenters physiques, les entreprises sont à présent en mesure d'agrandir ou de réduire de façon dynamique leurs data warehouses pour répondre rapidement à l'évolution des budgets et des exigences métiers.

Les architectures modernes dans le cloud allient trois éléments essentiels : la puissance du data warehouse, la flexibilité des plateformes de Big Data et l'élasticité du cloud, le tout à une fraction du prix des solutions traditionnelles. Cet e-book présente les meilleurs data warehouses dans le cloud en soulignant les différences notables et en s'appuyant sur une approche éprouvée pour faciliter l'accès aux données à tous les utilisateurs et maximiser la création de valeur.

## **TRAITEMENT MASSIVEMENT PARALLÈLE (MPP ; MASSIVELY PARALLEL PROCESSING)**

Les data warehouses qui prennent en charge les projets de Big Data utilisent des architectures de traitement massivement parallèle (MPP) pour fournir des requêtes à hautes performances sur d'importants volumes de données. Les architectures MPP sont constituées de nombreux serveurs qui s'exécutent en parallèle pour distribuer les charges de traitement et d'entrée/sortie (E/S).

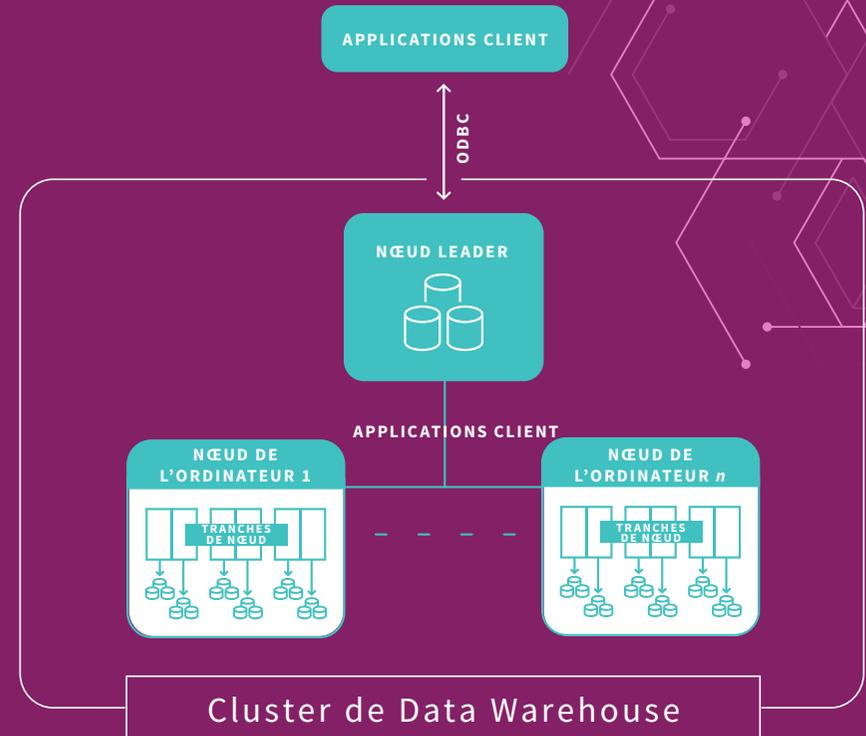
## **MAGASINS DE DONNÉES EN COLONNES**

Les data warehouses MPP sont généralement conçus comme des magasins de données en colonnes, car c'est la solution la plus flexible et la plus économique pour l'analytics. Ces bases de données stockent et traitent les informations par colonne plutôt que par ligne et accélèrent de façon spectaculaire les requêtes groupées, qui sont souvent utilisées pour le reporting.

# Amazon Redshift

Redshift est un service de data warehouse dans le cloud, entièrement géré, capable de traiter des pétaoctets de données. Commencez avec quelques gigaoctets de données avant d'évoluer jusqu'aux pétaoctets. De cette manière, vous pourrez acquérir de nouvelles informations dérivées de vos données métiers et clients.

La première étape du processus de création d'un data warehouse Redshift consiste à lancer un jeu de nœuds, appelé un cluster Amazon Redshift. Après avoir mis votre cluster en service, vous téléchargez votre jeu de données, puis lancez des requêtes d'analyse de données. Quelle que soit la taille de votre jeu de données, Amazon Redshift permet d'exécuter les requêtes avec une grande rapidité via des outils familiers basés sur SQL et des applications de Business Intelligence.



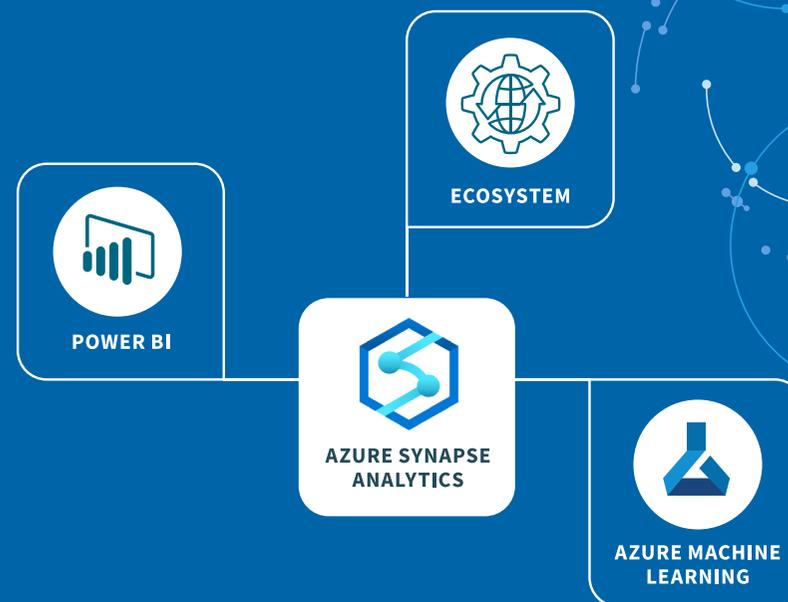
## LE DATA WAREHOUSE DANS LE CLOUD LE PLUS LARGEMENT ADOPTÉ

Pendant de nombreuses années, les data warehouses étaient uniquement disponibles en tant que solution sur site. Puis, en novembre 2012, Amazon Web Services (AWS) a lancé Redshift. Bien que plusieurs data warehouse dans le cloud l'aient précédé, il a été le premier à gagner des parts de marché via l'adoption. Le dialecte SQL de Redshift est basé sur PostgreSQL, qui est bien compris des analystes du monde entier, et s'appuie sur une architecture connue de beaucoup d'utilisateurs de data warehouses sur site.

# Microsoft Azure Synapse Analytics

Azure Synapse Analytics est un service d'analytics plus récent qui associe les Data Warehouses d'entreprise et l'analyse du Big Data. Ce système vous donne la liberté d'interroger les données via des ressources à la demande sans serveur ou via des ressources provisionnées. Azure Synapse offre une expérience unifiée pour ingérer, préparer, gérer et diffuser des données pour répondre à vos besoins en matière de Business Intelligence (BI) et de Machine Learning (ML).

Au cœur d'Azure Synapse réside un moteur de traitement SQL distribué, natif cloud. Ce service s'appuie sur SQL Server pour gérer vos charges de travail de data warehouse d'entreprise les plus exigeantes. Semblable à d'autres solutions MPP dans le cloud, Azure SQL Data Warehouse (SQL DW) sépare le stockage et le calcul, en facturant séparément chaque service. Azure Synapse enregistre les données des tables relationnelles via un stockage en colonnes et permet de se passer de machines physiques en représentant la puissance de calcul sous la forme d'unités de data warehouse (DWU). De cette façon, vos utilisateurs peuvent faire évoluer les ressources de calcul facilement, de façon transparente et à volonté.



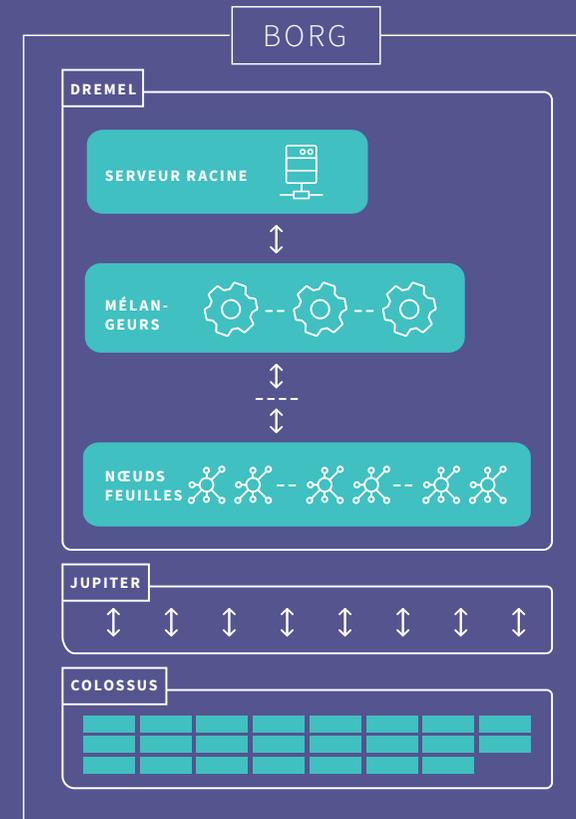
## EXPLOITER SQL AU-DELÀ DU DATA WAREHOUSE

Synapse Analytics vise à unifier tout un éventail de charges de travail d'analyse, comme les data warehouses, les data lakes et le Machine Learning en une interface utilisateur unique. La combinaison d'un moteur SQL, d'Apache Spark avec Azure Data Lake Storage (ADLS) et Azure Data Factory donne aux utilisateurs la possibilité de contrôler à la fois les data warehouses/data lakes et la préparation des données pour les tâches de Machine Learning. Azure Synapse facilite une évolutivité verticale et horizontale du data warehouse. Verticale, car il modifie le niveau de service ou place la base de données dans un pool élastique. Horizontale, car il ajoute davantage d'unités de data warehouse.

# Google BigQuery

BigQuery est un data warehouse entièrement géré, sans serveur, qui évolue automatiquement pour répondre aux besoins en matière de stockage et de puissance de calcul. Avec BigQuery, vous obtenez une base de données en colonnes et ANSI SQL capable d'analyser des téraoctets ou même des pétaoctets de données à des vitesses fulgurantes. BigQuery vous permet également de réaliser des analyses de données géospatiales en combinant un système SQL qui vous est familier avec un système d'information géographique BigQuery. En outre, vous pouvez rapidement créer et appliquer des modèles de Machine Learning sur des données à grande échelle structurées ou semi-structurées en combinant un système SQL simple avec le Machine Learning de BigQuery. Vous pouvez également prendre en charge la création interactive et en temps réel de tableaux de bord avec le moteur BI de BigQuery.

L'architecture BigQuery est constituée de plusieurs composants. Borg est le système de calcul. Colossus, le stockage distribué. Jupiter, le réseau. Et Dremel, le moteur d'exécution.



## UNE SOLUTION SANS SERVEUR

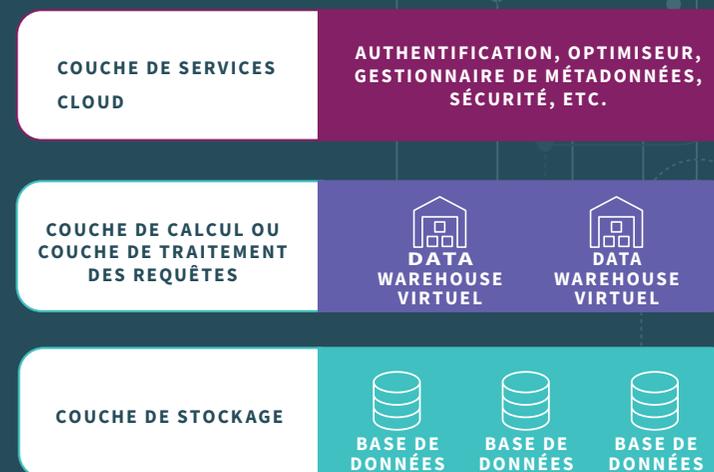
Google n'attend pas de vous que vous assuriez la gestion de l'infrastructure de votre data warehouse, c'est pourquoi BigQuery masque un grand nombre de détails sous-jacents inhérents au matériel, à la base de données, aux nœuds et à la configuration. Son élasticité en fait une solution automatiquement prête à l'emploi. Pour démarrer, il suffit simplement de créer un compte sur Google Cloud Platform (GCP), de charger une table et d'exécuter une requête. Google s'occupe du reste.

# Snowflake Cloud Data Platform

Snowflake est un data warehouse MPP dans le cloud, entièrement géré, qui s'exécute sur AWS, GCP et Azure. Si vous êtes un utilisateur de Snowflake, vous pouvez sélectionner autant de data warehouses virtuels que nécessaire pour mettre en parallèle et isoler les performances de requêtes individuelles. Snowflake autorise une simultanée très élevée en séparant le stockage et le calcul, afin de garantir que de nombreux data warehouses puissent accéder simultanément à la même source de données.

Vos interactions avec le data warehouse de Snowflake s'effectuent par l'intermédiaire d'un navigateur Web, de la ligne de commande, d'une plateforme d'analyse, ou du pilote ODBC ou JDBC de Snowflake, ou encore d'autres pilotes pris en charge. La plateforme prend en charge le traitement relationnel compatible avec ACID et dispose d'une prise en charge native pour les formats de stockage de documents tels que JSON, Avro, ORC (Optimized Row Columnar), Parquet et XML.

L'architecture hybride de Snowflake est divisée en trois couches distinctes :



## LE PREMIER DATA WAREHOUSE MULTI-CLOUD

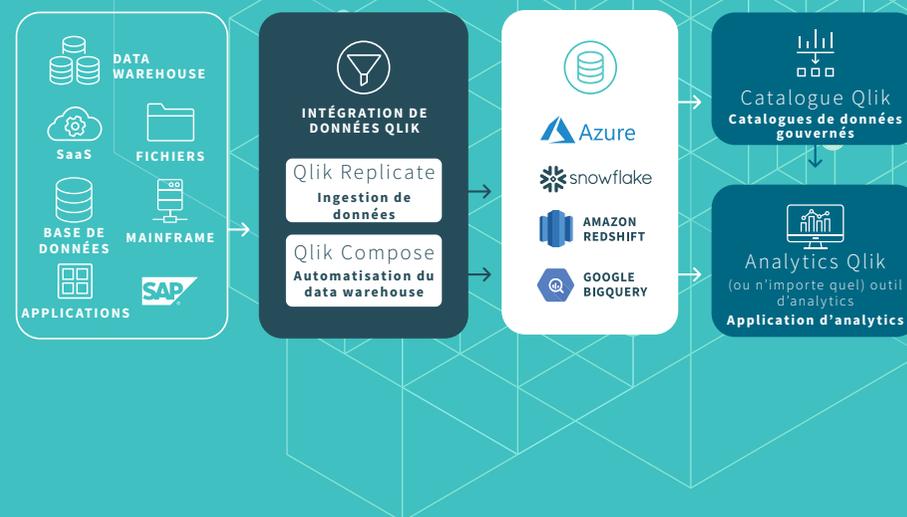
Contrairement aux autres data warehouses dont nous avons dressé le profil, Snowflake est la seule solution qui ne s'exécute pas sur son propre cloud. Il s'agit du premier data warehouse multi-cloud disponible sur AWS, GCP et Azure. Avec une base de code commune et interchangeable, Snowflake est doté d'une fonctionnalité de réplication de données globale, ce qui signifie que vous pouvez transférer vos données dans n'importe quel cloud et n'importe quelle région, sans avoir à réécrire le code de vos applications ou à acquérir de nouvelles compétences.

# Rapide aperçu des meilleurs Data Warehouses dans le Cloud

	Amazon Redshift	Microsoft Azure Synapse	Google BigQuery	Snowflake Cloud Data Platform
Version initiale	2012	2016	2010	2014
Sépare le stockage et le calcul	Non	Oui	Oui	Oui
Multi-cloud	Non	Non	Non	Oui
Langage de requête	SQL Amazon Redshift	TSQL	SQL 2011 standard & SQL BigQuery	SQL Snowflake
Élasticité	Oui - Manuelle	Oui – Manuelle et automatique	Oui – Automatique	Oui – Automatique
MPP	Oui	Oui	Oui	Oui
En colonnes	Oui	Oui	Oui	Oui
Clés étrangères	Oui	Oui	Non	Oui
Transaction	ACID	ACID	ACID	ACID
Simultanéité	Oui	Oui	Oui	Oui
Durabilité	Oui	Oui	Oui	Oui
Automatisation	Non	Non	Non	Non
Site Web	Lien	Lien	Lien	Lien
Essai gratuit	Oui	Oui	Oui	Oui

# Un Data Warehouse agile

Notre plateforme Qlik® Data Integration Platform (anciennement Attunity) automatise l'ensemble du cycle de vie du data warehouse afin d'accélérer la disponibilité de vos données prêtes à l'emploi. Notre approche axée sur les modèles aide vos data engineers à concevoir, déployer, gérer et cataloguer des data warehouses sur mesure dans le cloud plus rapidement que les solutions classiques. Ajoutez Qlik à tout data warehouse dans le cloud de votre choix et concrétisez les promesses du data warehouse agile en termes de coût et d'efficacité.



## NOS FACTEURS DE PRODUCTIVITÉ QLIK

- **Ingestion et mises à jour de données en temps réel** – Une solution simple et universelle permettant une ingestion continue et en temps réel de vos données d'entreprise au sein des data warehouses dans le cloud les plus répandus.
- **Flux de travail automatisé** – Une approche axée sur les modèles permettant d'affiner continuellement vos opérations de data warehouse.
- **Des données fiables, adaptées à l'entreprise** – Un catalogue de données intelligent, à l'échelle de l'entreprise, permettant de partager en toute sécurité vos data marts.

# Optez pour un Data Warehouse dans le Cloud et innovez avec Qlik

Le cloud est désormais la plateforme incontournable de l'analytics moderne. C'est la raison pour laquelle votre entreprise a besoin d'approches et de technologies qui simplifient l'analytics dans le cloud, grâce à une solution qui produit davantage de valeur avec des itérations plus rapides et moins de ressources. Avec Qlik vous automatisez votre data warehouse, optimisez votre pipeline de données, fournissez un catalogue de données sécurisé et, comme si cela ne suffisait pas, vous bénéficiez des meilleures fonctionnalités d'analytics de l'industrie.

Qlik représente une solution robuste, complète et novatrice pour la mise en œuvre d'une architecture, d'une intégration, d'une fourniture et d'une analyse de données de tout premier plan pour votre entreprise.

Vous êtes prêt pour le data warehouse agile ? Nous sommes là pour vous aider.

Pour de plus amples informations, allez sur [qlik.com/fr-fr/data-warehouse-automation](http://qlik.com/fr-fr/data-warehouse-automation).



## À PROPOS DE QLIK

---

Qlik s'est donné pour objectif la création d'un monde « data literate », où chacun peut exploiter les données pour résoudre les défis les plus complexes. Seul Qlik est à même de proposer des solutions d'analyse et d'intégration de données en temps réel et de bout en bout qui aident les organisations à transformer la data en valeur. Qlik aide les entreprises à piloter par la donnée afin de mieux comprendre le comportement de leurs clients, réinventer les processus métier, découvrir de nouvelles sources de revenus et équilibrer le rapport bénéfices/risques. Qlik exerce ses activités dans plus de 100 pays et offre ses services à plus de 50 000 clients à travers le monde.

