



## Sommaire



### ***Les données sensibles dans le Cloud***

- ❖ *Plus les données sont sensibles plus elles vont dans le Cloud*
- ❖ *Une étape symbolique de la pérennisation du Cloud*
- ❖ *Séparer le stockage des données de leur traitement*
- ❖ *DaaS : convergence des données monoclient, comment et pourquoi*
- ❖ *Fédération ouverte multi Cloud : comment et pourquoi*
- ❖ *L'extension dynamique de l'entrepôt par le fédérateur*
- ❖ *Snowflake, un entrepôt en mode fédéré*
- ❖ *Avantages et inconvénients de Snowflake, les risques pris*
- ❖ *Un écosystème nouveau apparaît*

Selon PRNewswire : 10,4 G\$ en 2026 (CAGR de 22,56 %) et 39,1 G\$ pour Industry arc. MarketsandMarkets est à 12,9 G\$.

# Plus les données sont sensibles

## ...plus elles vont dans le Cloud

- ❖ La crainte des usagers est de confier à des prestataires externes étrangers, des données sensibles et vitales pour l'entreprise.
- ❖ On peut aussi raisonner à l'inverse.
- ❖ Nous pratiquons un métier qui n'est pas le nôtre :
  - ❖ Démarrer les serveurs
  - ❖ Réactiver un switch
  - ❖ Rebuter un routeur
  - ❖ Décrire le schéma logique d'une base de données...
- ❖ Mais cela fait 50 ans que cela dure et il n'y a pas de raison que cela s'arrête.
- ❖ A terme nous ne serons plus concernés par les aspects matériels et logiciels des infrastructures, fussent-elles chargées d'héberger nos données sensibles.
- ❖ Si l'entreprise doit communiquer (hors militaires et organisations secrètes)
  - ❖ La meilleure manière de garantir la qualité et la disponibilité des données sensibles est de les confier à ceux qui en ont la pratique.
  - ❖ Avec toutes les précautions qui les accompagnent : juridiques, sauvegardes, tests en vraie grandeur de récupération, groupement d'utilisateurs pour "surveiller" la politique des prestataires...
- ❖ Les questions à se poser :
  - ❖ Qui est capable de mettre les ressources nécessaires à la gestion des infrastructures ?
  - ❖ Qui dispose des personnels système et sécurité pour réagir en cas de panne ou d'attaque ?
- ❖ Sur le long terme, à partir de 2030, ON N'AURA PAS LE CHOIX.
- ❖ D'où problèmes de dépendance.
- ❖ Il n'y a pas de solution miracle.



## De quelles données parle-t-on ?

- ❖ Tout ce qui a un caractère de sensibilité, risques d'intégrité, confidentialité, données personnelles...
- ❖ Difficile d'établir les frontières
  - ❖ Données de santé
  - ❖ Données financières de l'entreprise
  - ❖ Données judiciaires
  - ❖ Données personnelles : identité, goûts, choix effectués...
  - ❖ Données clients
  - ❖ etc
- ❖ Comment respecter les contraintes liées à ces données
  - ❖ Les conserver en interne : a-t-on les moyens de les protéger
  - ❖ Les mettre en lieu sûr ? Le Cloud ?
  - ❖ Qui a le plus à perdre, le prestataire ou le client ?
- ❖ Les blocages sont historiques, ils ne correspondent plus à la réalité



# Une étape dans la pérennisation du Cloud

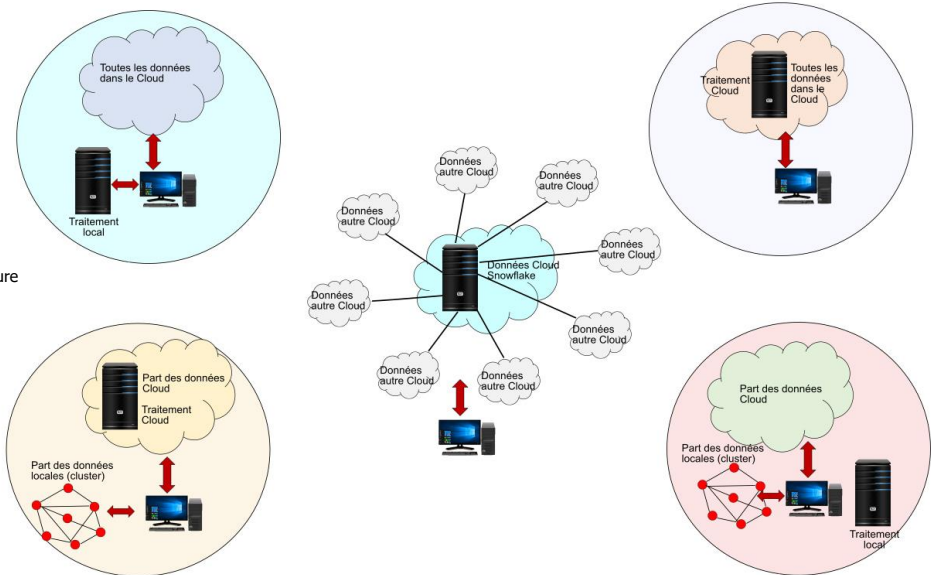
- ❖ Le "danger" que représentent les prestataires et les GAFAM inquiètent moins :
  - ❖ Ils sont dans le paysage
  - ❖ On s'est habitués
  - ❖ Il n'y a pas de gros scandales de pertes de données liés spécifiquement au Cloud... pour l'instant
- ❖ Les avantages sont indéniables
  - ❖ Une disponibilité permanente
  - ❖ Des coûts réduits plus faciles à maîtriser
  - ❖ Des fragilités d'infrastructures qui ne nous concernent plus
  - ❖ Une plus grande réactivité (résilience) par rapport aux nouveaux besoins, qui sont toujours à mettre en place pour le lendemain matin...
  - ❖ Une plus grande ouverture aux formats
  - ❖ Un respect incontestable des contraintes règlementaires, européennes (RGPD) ou américaines
  - ❖ Des performances garanties, grâce au respect des fameux SLA
- ❖ Le basculement des traitements est confirmé, celui des données sensibles va suivre



Les comptables ont longtemps refusé la dématérialisation. Ce n'est plus le cas aujourd'hui, même si certains continuent d'imprimer six fois les factures numériques. Il en sera de même pour les données sensibles.

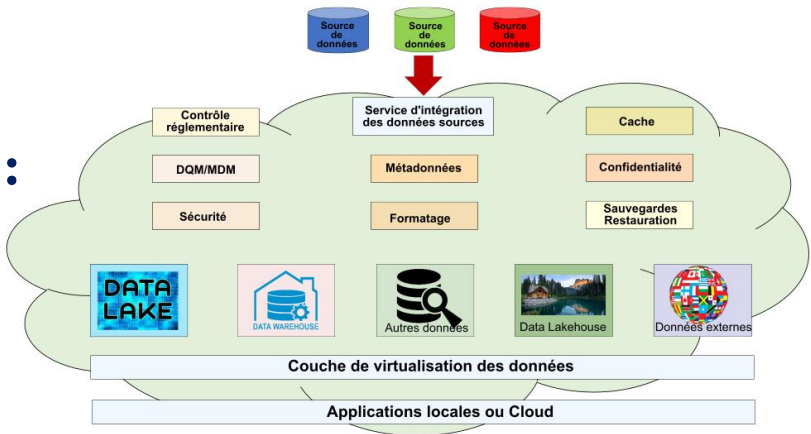
## Séparer le stockage du traitement

- ❖ L'idée du DaaS est de séparer le stockage des données de leur traitement.
- ❖ On peut imaginer une infrastructure avec des données entièrement portées par un Cloud et les traitements restés en local



# DaaS standard : convergence des données

## Comment et pourquoi



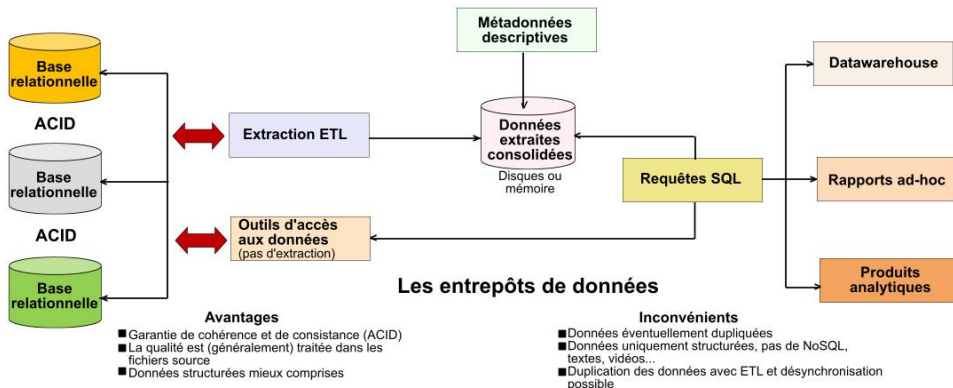
Le DaaS est une sorte de disque virtuel géré par un prestataire externe



Les données sensibles dans le Cloud

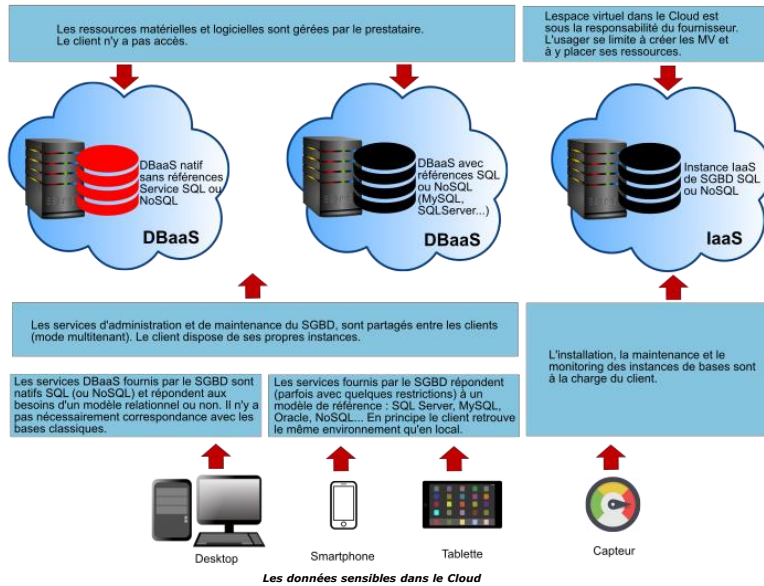
# Le cas particulier des entrepôts de données

## Le mode classique du datawarehouse relationnel



Les données sensibles dans le Cloud

# Le même problème que les bases de données

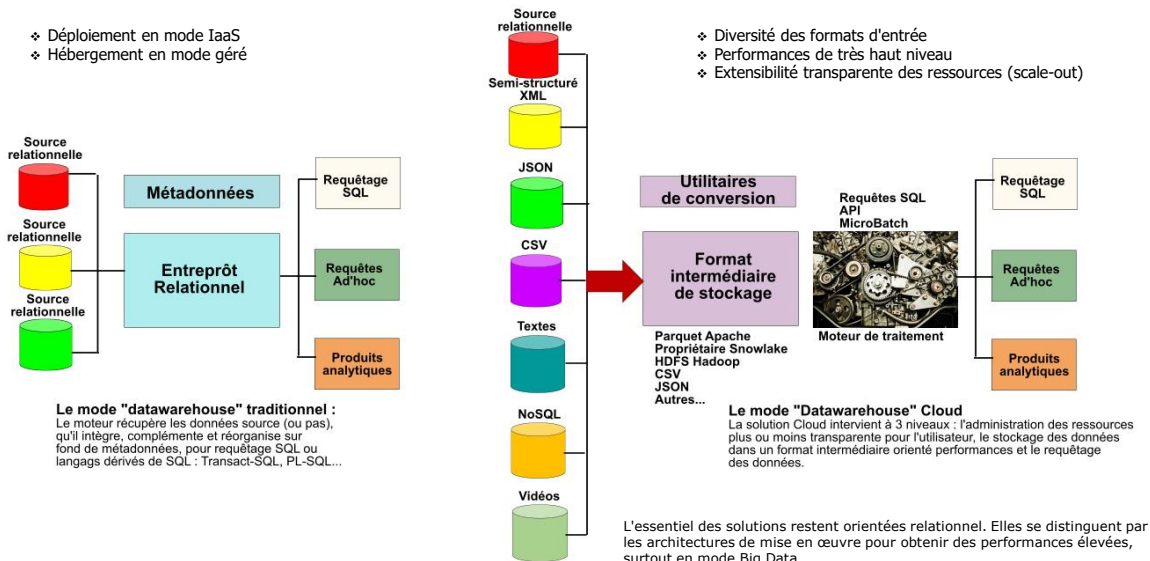


9 / 15

## Le rôle du Cloud

- ❖ Déploiement en mode IaaS
- ❖ Hébergement en mode géré

- ❖ Diversité des formats d'entrée
- ❖ Performances de très haut niveau
- ❖ Extensibilité transparente des ressources (scale-out)



Les données sensibles dans le Cloud

10 / 15

# L'extension dynamique de l'entrepôt

- ❖ SnowFlake est une solution française d'origine (2012) d'entrepôt SaaS (géré)
- ❖ Dispose de son propre Cloud, mais aussi sur AWS (7 régions), GCP, Azure (6 régions)... et Alibaba
- ❖ En fait un Cloud mondial (13 régions) ouvert aux principaux Clouds publics
- ❖ Les fondateurs de SnowFlake sont d'origine Oracle : ils veulent être plus rapides et moins chers qu'Oracle
- ❖ SnowFlake est structuré autour d'un cœur relationnel accessible en SQL, conçu pour le Cloud et pas l'inverse
- ❖ Micro-partitionnement par lignes et stockage en mode colonnes (comme une structure NoSQL Cassandra ou Big Table de Google)
- ❖ Indépendance complète entre le traitement et le stockage
- ❖ Les traitements sont effectués par des clusters dédiés : un par entrepôt virtuel
- ❖ Ce n'est pas un ETL, mais supporte n'importe quel outil d'intégration compatible JDBC, ODBC
- ❖ Ce n'est pas un outil d'analyse, mais il supporte la grande majorité des outils spécialisés
- ❖ Les connexions applicatives :
  - ❖ SnowSQL (client CLI architecture .NET)
  - ❖ SnowFlake for Python
  - ❖ SnowFlake for Spark
  - ❖ SnowFlake for Kafka
  - ❖ Node.js driver
  - ❖ Go SnowFlake driver
  - ❖ .NET driver
  - ❖ ODBC driver
  - ❖ PHP PDO Driver for SnowFlake
  - ❖ SnowFlake SQL API
- ❖ Ne peut pas être installé en local, ni en Cloud privé dans un Cloud public



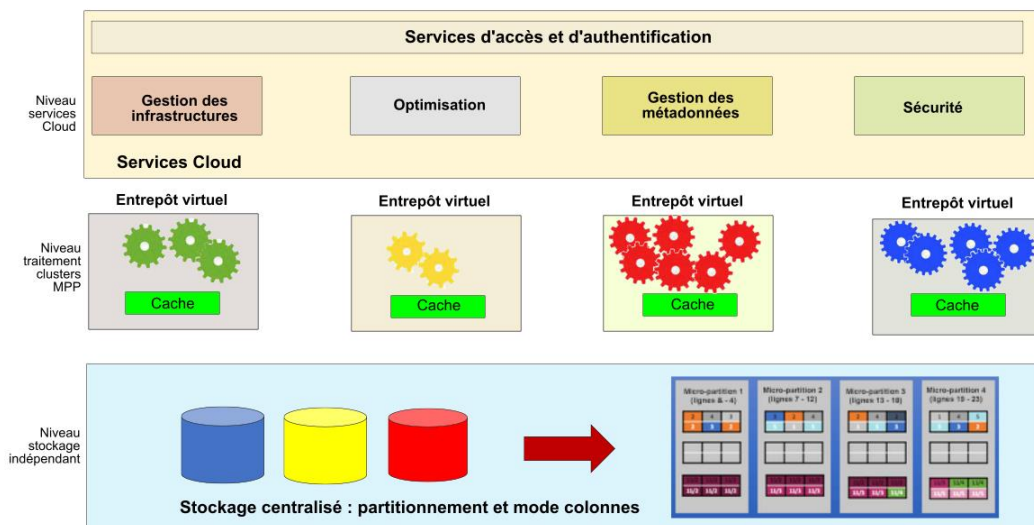
Une nouvelle génération d'entrepôts de données

Snowflake Data Platform?

5 milliards de requêtes attendues en 2022  
 2.500 clients (500 de plus chaque trimestre)  
 A la fin 2020, SnowFlake était estimé à 70 G\$  
 Mais son modèle de licence est-il crédible ?

## SnowFlake : fédération ouverte multi Cloud :

SnowFlake est un cas particulier du Datawarehouse dans le Cloud



# Snowflake

## Avantages et inconvénients, les risques pris

### Avantages

- ❖ Tout est géré par Snowflake : matériel, logiciel, mises à jour
- ❖ Séparation entre le stockage et les traitements
- ❖ Optimisation automatique : indexation, partitionnement
- ❖ Constitution de clones sans duplication des données
- ❖ Adéquation des ressources en mode scale-out
- ❖ Intégration de données issues d'autres Clouds sans copier les données
- ❖ Performances très élevées sans les contraintes
- ❖ Paiement à la consommation : attention aux coûts
- ❖ Accès facile aux versions antérieures de données
- ❖ Grande diversité d'import de données : Fichiers plats CSV, fichiers JSON, Avro, ORC, Parquet et XML, fichiers S3, etc.
- ❖ Snowflake s'appuie sur Matillion, un partenaire, qui se charge de toutes les opérations d'intégration de données issues des Clouds multiples

### Inconvénients

- ❖ Le format propriétaire du Cloud peut ajouter à la complexité et pénaliser les performances
- ❖ Pas de communauté Open Source ouverte pour faire avancer la technologie (syndrome Docker) : le plus gros défaut, Snowflake n'est pas en phase avec l'histoire...
- ❖ Les procédures stockées sont écrites en JavaScript, script Snowflake et Scala, des langages pas forcément pratiqués par les usagers



## Un écosystème nouveau apparaît





# Les données sensibles dans le Cloud

10 Juin 2022

## Nos prochains webinaires

23 Juin 2022 :

1 Juillet 2022 :

2 Septembre 2022 :

L'assistance "intelligente" au codage

La programmation du comportement des réseaux

Le "tout en un" de l'hyperconvergence



[claudio@lemarson.com](mailto:claudio@lemarson.com)

<https://www.lemarson.com>

Les données sensibles dans le Cloud

15 / 15