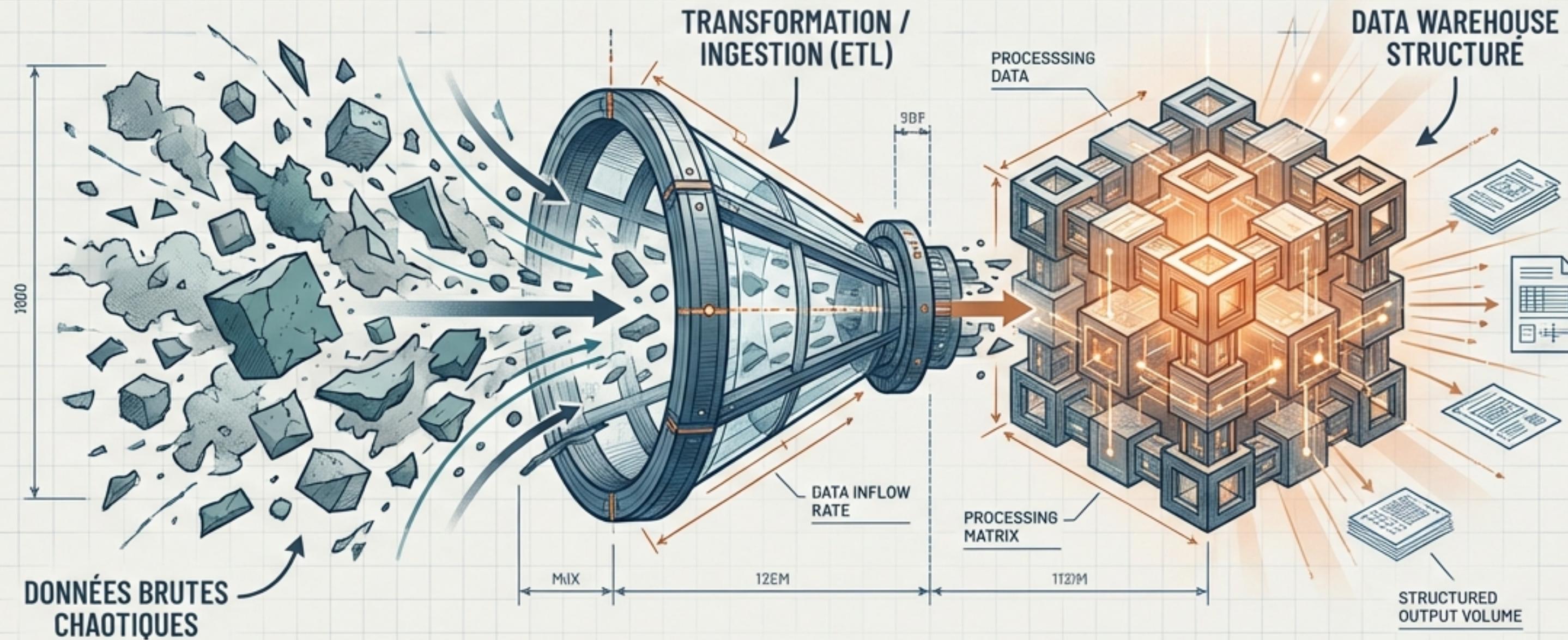


# INTRODUCTION AU DATA WAREHOUSE

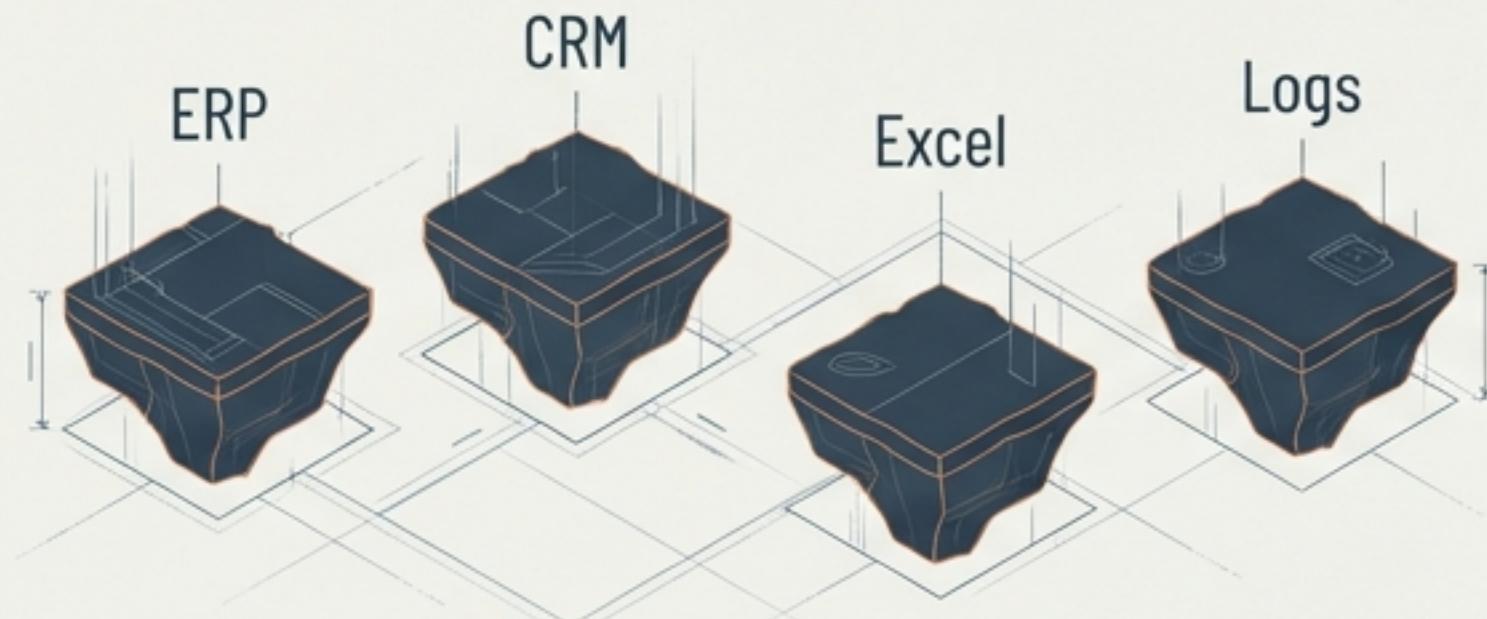
Architecture, Concepts et Évolution



# DU CHAOS DES SILOS À LA CLARTÉ CENTRALISÉE

Le problème de la donnée dispersée

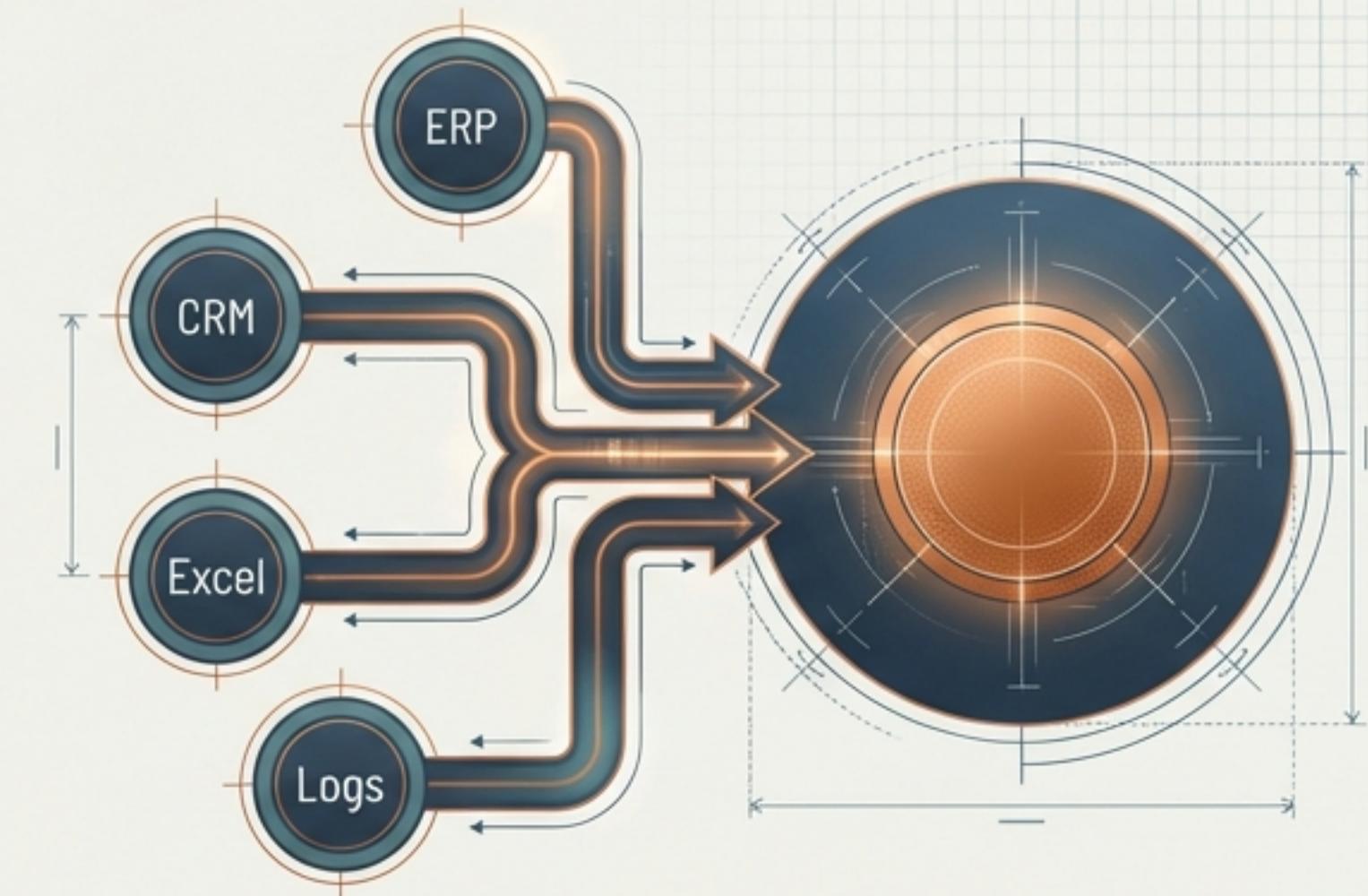
## THE PROBLEM



### Les Douleurs :

- **Incohérence** : Données dispersées et difficiles à réconcilier.
- **Visibilité limitée** : Impossible de croiser les informations pour une vue consolidée.
- **Performance** : Les requêtes analytiques lourdes ralentissent les systèmes opérationnels.

## THE SOLUTION



La centralisation est la seule voie vers une prise de décision fiable.

# QU'EST-CE QU'UN DATA WAREHOUSE ?

**Un Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour supporter le processus de prise de décision.**

– Bill Inmon (1992)

## Concept Clé

Un système de stockage centralisé conçu spécifiquement pour l'analyse et le reporting, distinct des bases de données transactionnelles.



Consolidation de multiples sources → Décision éclairée



# LES 4 CARACTÉRISTIQUES FONDAMENTALES



## 1. Orienté sujet

Organisé autour des thèmes majeurs de l'entreprise.

*Exemple : Ventes, Clients, Produits.*



## 2. Intégré

Uniformisation des données provenant de sources variées.

*Exemple : Standardisation des codes pays (FR vs France).*



## 3. Non volatile

Données en lecture seule. Une fois chargées, elles ne sont pas modifiées.

*Exemple : Préservation stricte de l'historique.*



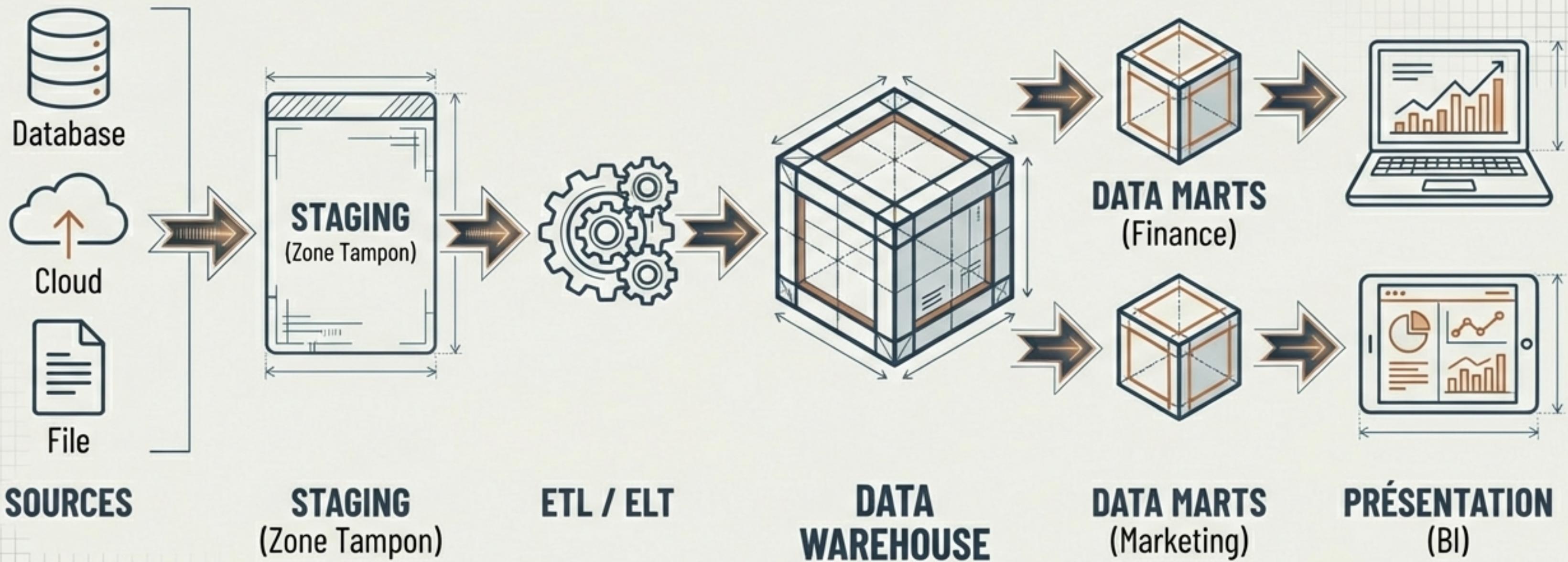
## 4. Historisé

Conservation de l'évolution temporelle des données.

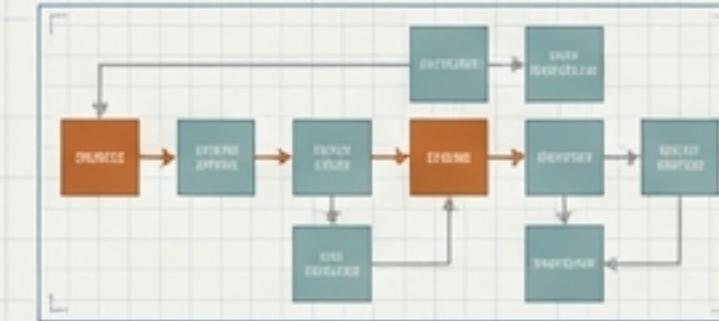
*Exemple : Analyse de l'évolution des prix au fil du temps.*

# L'ARCHITECTURE END-TO-END

Le flux de la donnée : de la source brute à l'insight métier

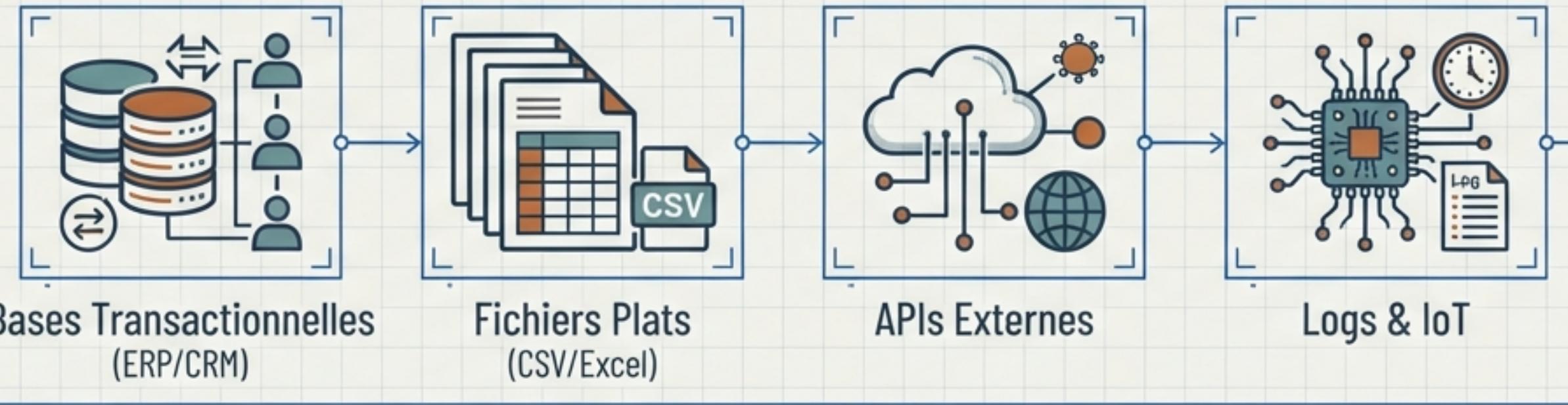


# INGESTION ET ZONE DE STAGING



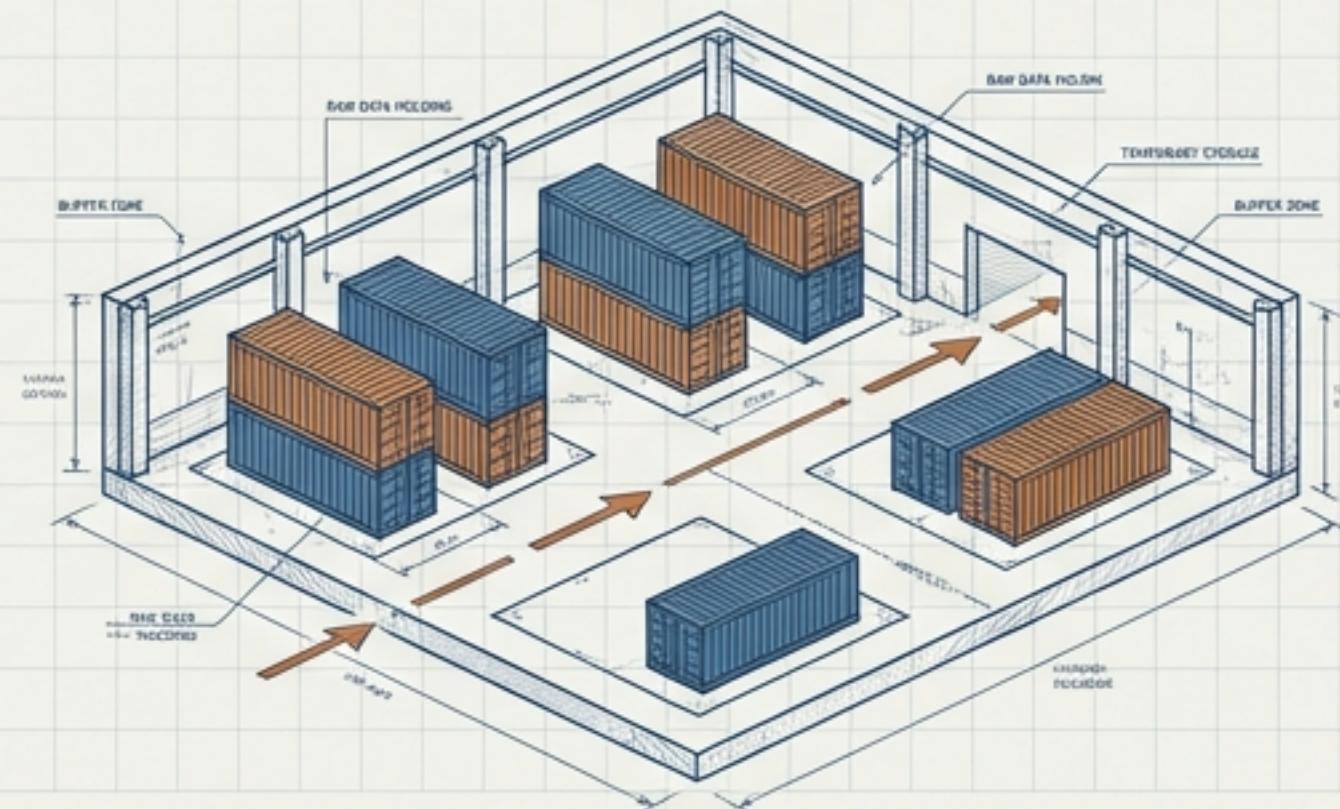
# LES SOURCES

## Les systèmes opérationnels qui alimentent l'entrepôt.

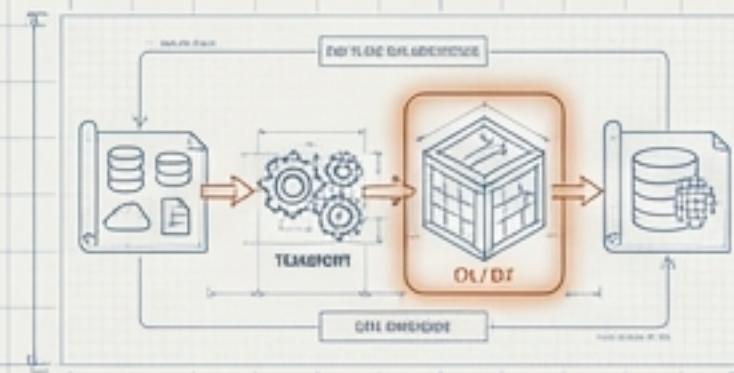


# LA ZONE DE STAGING (Landing Zone)

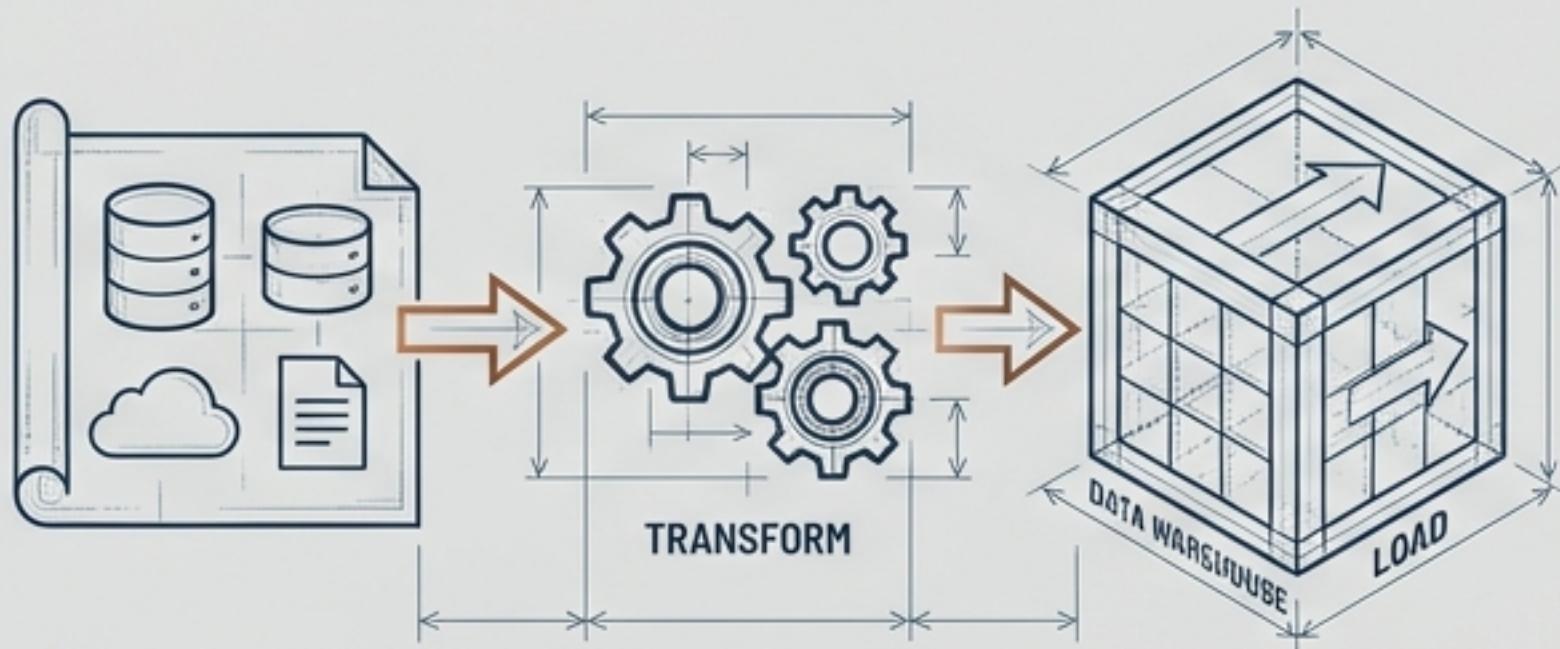
Une zone temporaire critique pour le chargement initial.  
Elle stocke les données brutes avant toute  
transformation complexe pour ne pas impacter les  
systèmes sources.



# LE MOTEUR DE TRANSFORMATION : ETL VS ELT



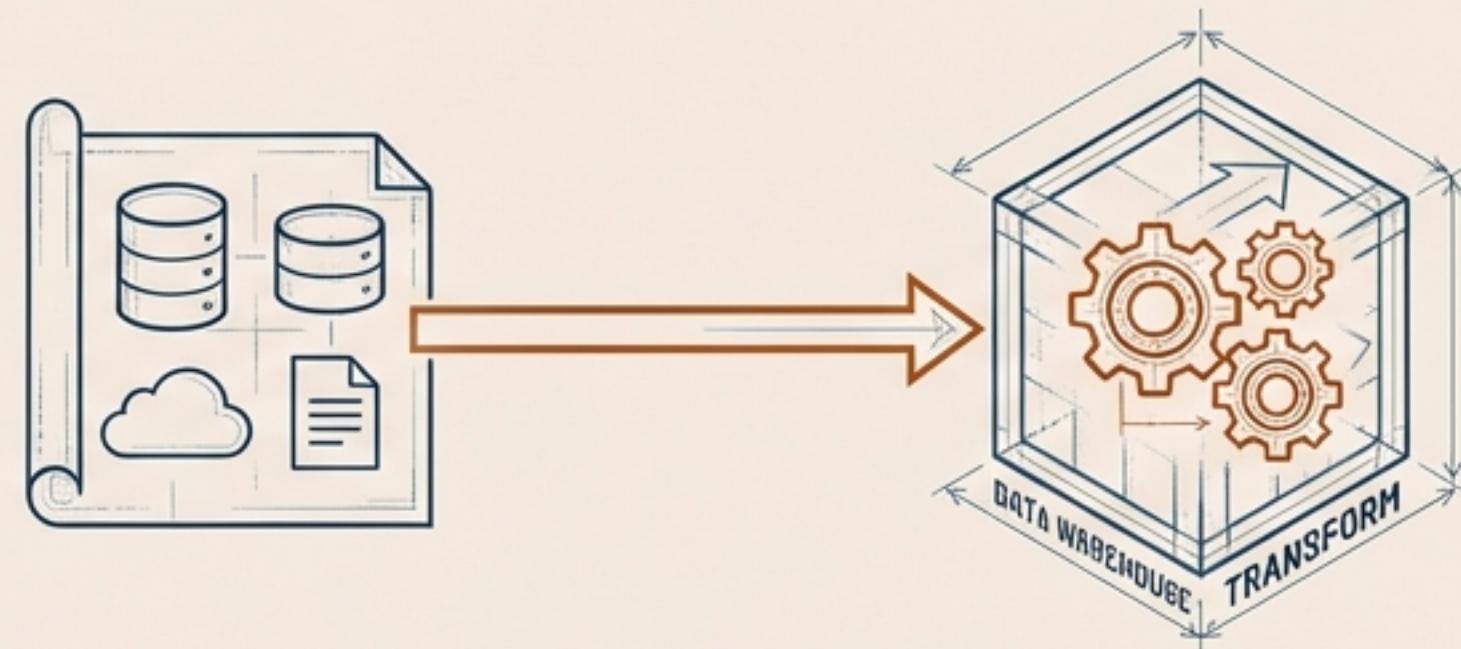
## ETL (Extract → Transform → Load)



**La donnée est transformée AVANT d'entrer dans le Warehouse.**

Approche traditionnelle, transformations complexes hors base.

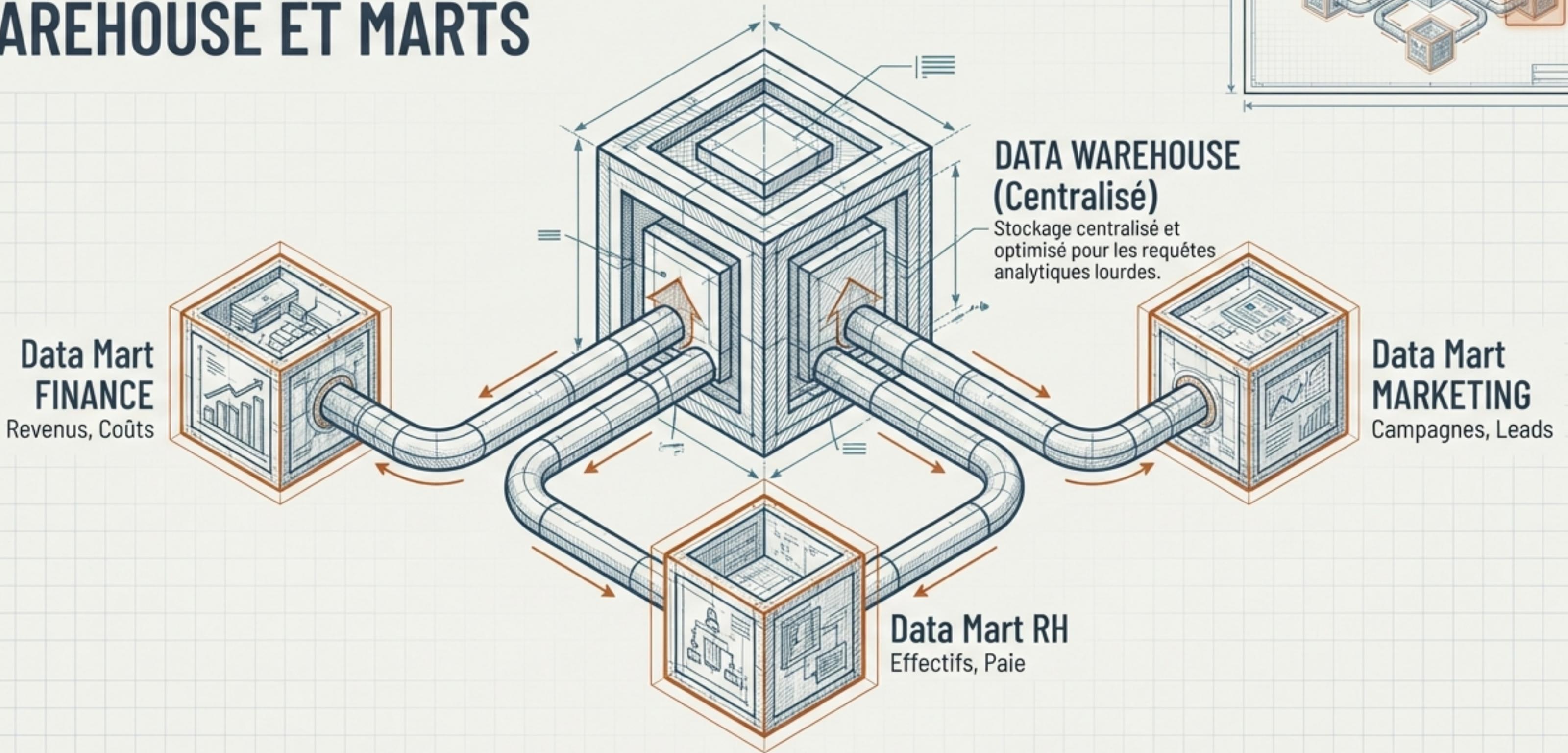
## ELT (Extract → Load → Transform)



**La donnée est chargée brute, puis transformée DANS le Warehouse.**

Standard du Cloud moderne (Snowflake, BigQuery). Utilise la puissance du DW.

# ORGANISATION DU STOCKAGE : WAREHOUSE ET MARTS



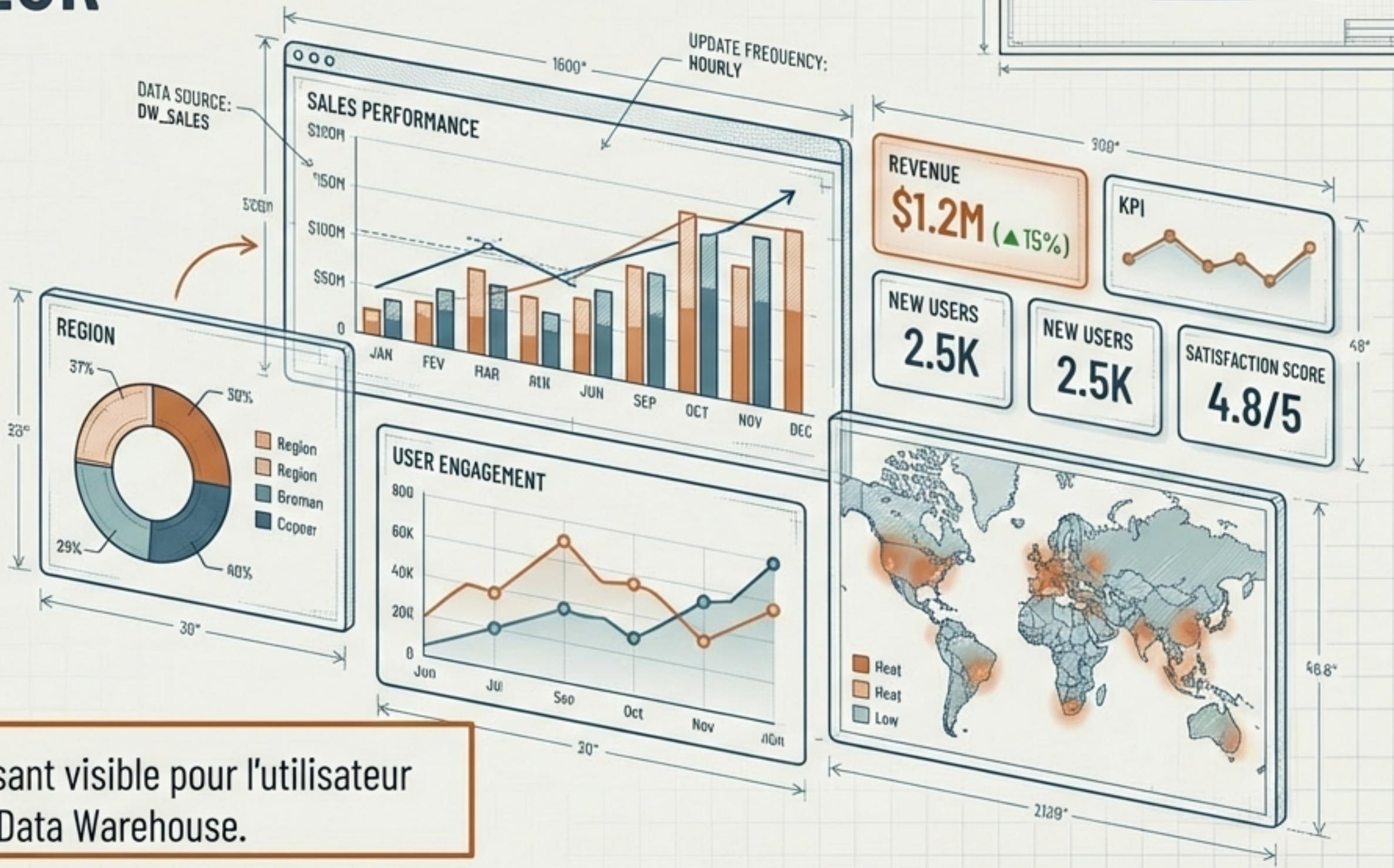
Les Data Marts sont des sous-ensembles du DW dédiés à un département spécifique.



# LA COUCHE DE PRÉSENTATION : RÉVÉLER LA VALEUR

Les outils de visualisation  
qui transforment les  
données en décisions.

- Outils populaires :  
Power BI, Tableau,  
Looker, Data Studio.



# CAS D'USAGE ET UTILISATEURS CIBLES

**BUSINESS INTELLIGENCE**



**Who:**  
Exécutifs & Managers

**What:**  
Tableaux de bord,  
Reporting  
mensuel/trimestriel,  
Suivi des KPIs.

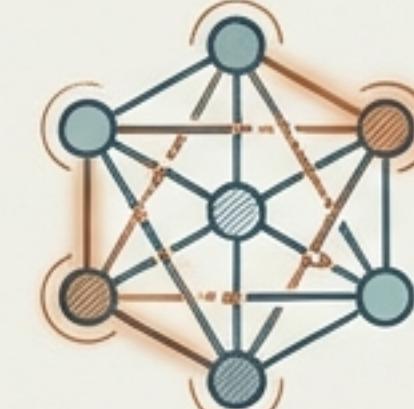
**ANALYSE AD-HOC**



**Who:**  
Analystes Métier

**What:**  
Exploration de données  
(Data Mining), questions  
ponctuelles,  
segmentation client.

**DATA SCIENCE**

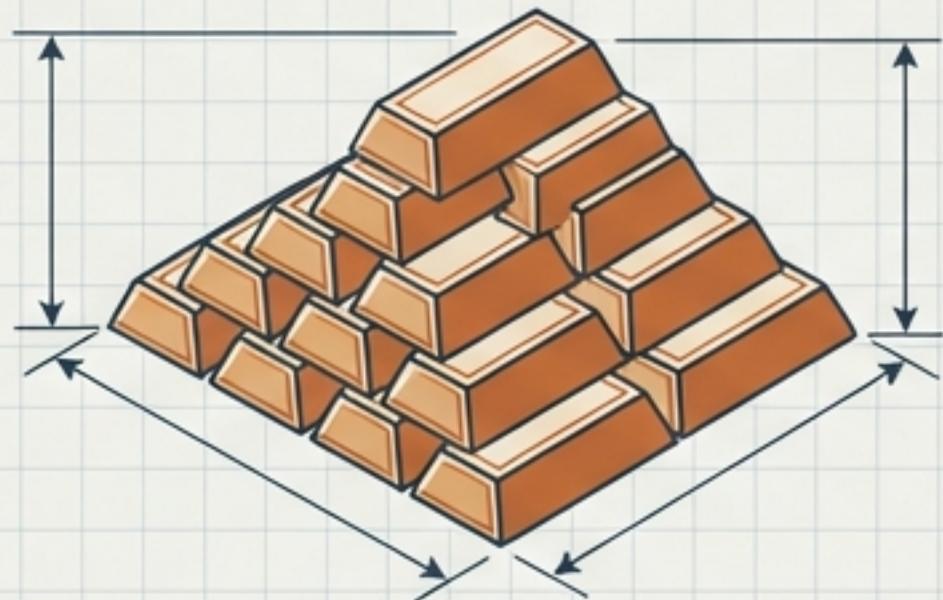


**Who:**  
Data Scientists

**What:**  
Préparation de datasets  
pour le Machine Learning,  
Feature engineering,  
Modélisation statistique.

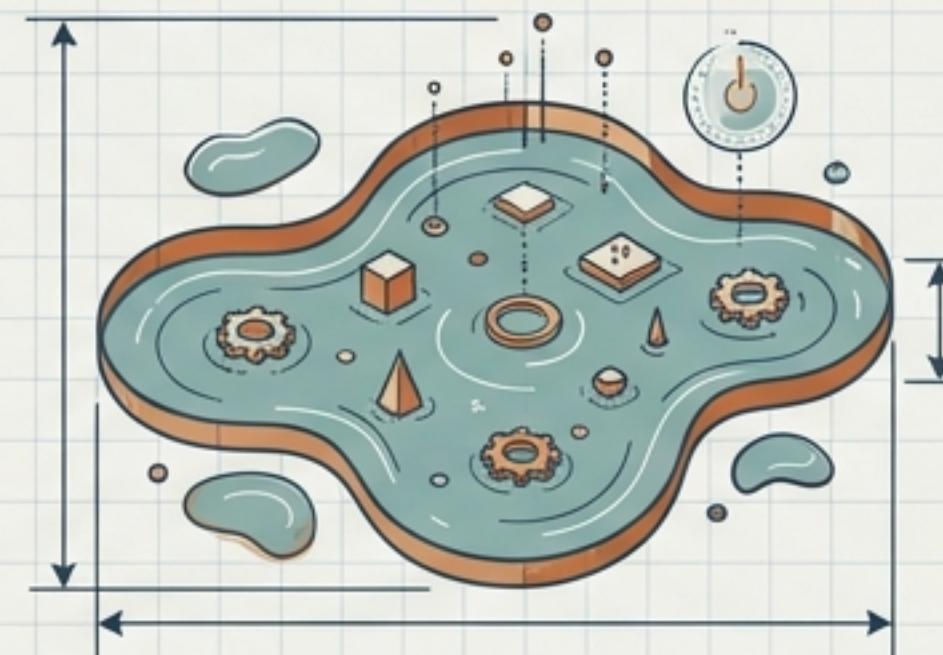
# COMPARATIF : DATA WAREHOUSE VS DATA LAKE

## DATA WAREHOUSE (L'Ordre)



- **Structure :** Données structurées & propres
- **Schema :** Schema-on-write (défini à l'écriture)
- **Qualité :** Haute (Curatée)
- **Coût :** Élevé (Stockage optimisé)
- **Utilisateurs :** Analystes, BI

## DATA LAKE (La Flexibilité)



- **Structure :** Toutes données (Brutes, non-structurées)
- **Schema :** Schema-on-read (défini à la lecture)
- **Qualité :** Variable
- **Coût :** Bas (Stockage brut)
- **Utilisateurs :** Data Scientists

# L'ÉVOLUTION : LE LAKEHOUSE



Le Lakehouse combine le meilleur des deux mondes.



Stockage économique  
(comme le Lake)



Performance et Gouvernance  
(comme le Warehouse)



Support unifié pour le  
SQL (BI) et Python (ML)

Exemples : Databricks, Delta Lake, Microsoft Fabric

# EN RÉSUMÉ : CE QU'IL FAUT RETENIR

- 1
- 2
- 3
- 4
- 5

-  Le **Data Warehouse** centralise les données hétérogènes pour permettre l'**analyse**.
-  Les **4 piliers** : Orienté sujet, Intégré, Non volatile, Historisé.
-  L'architecture moderne privilégie l'**ELT (Extract-Load-Transform)** dans le **Cloud**.
-  Les **Data Marts** offrent des **vues spécifiques** par métier.
-  L'avenir est au **Lakehouse** (**Convergence Warehouse + Lake**).

End of Module 01