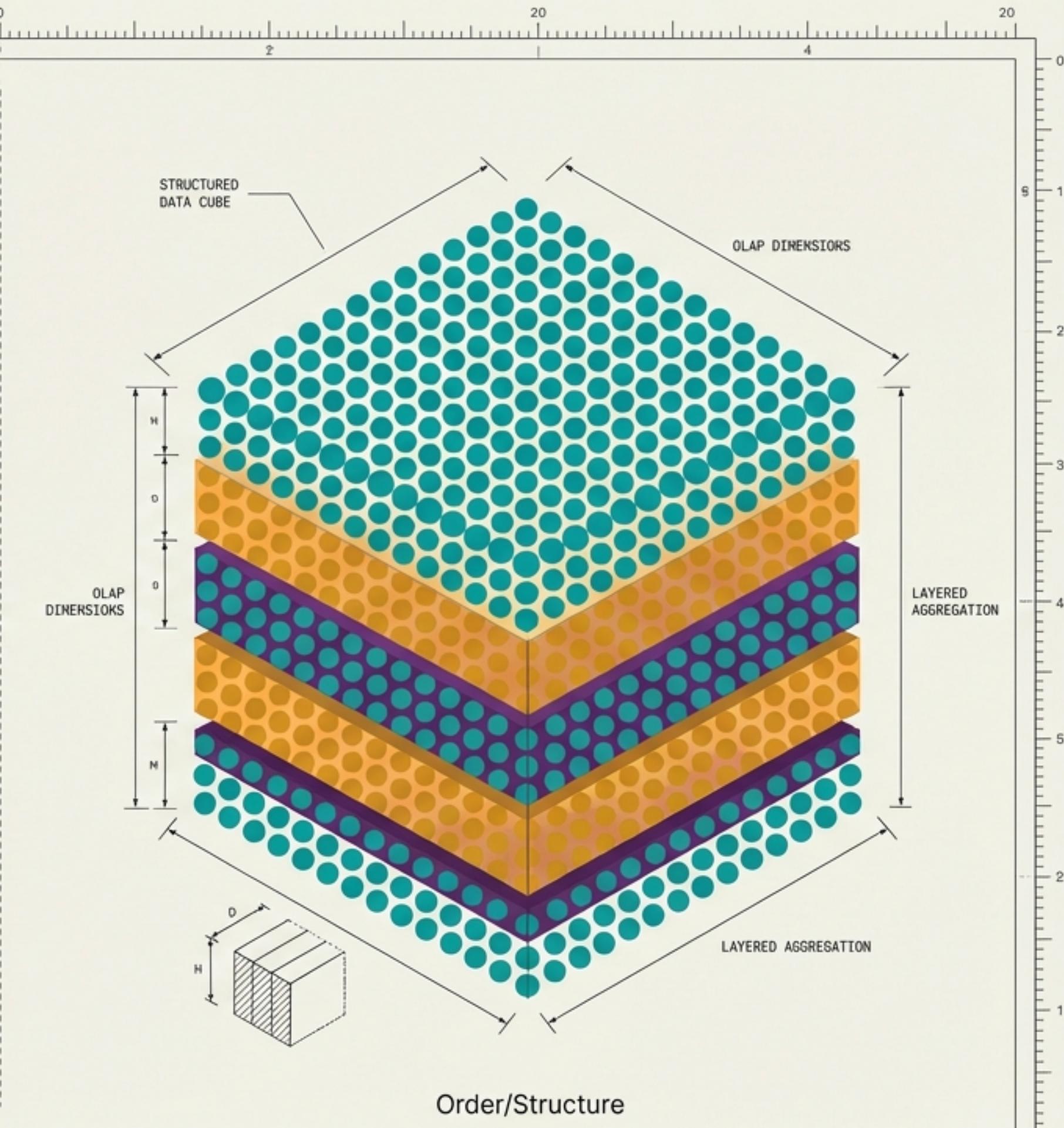


Module 02

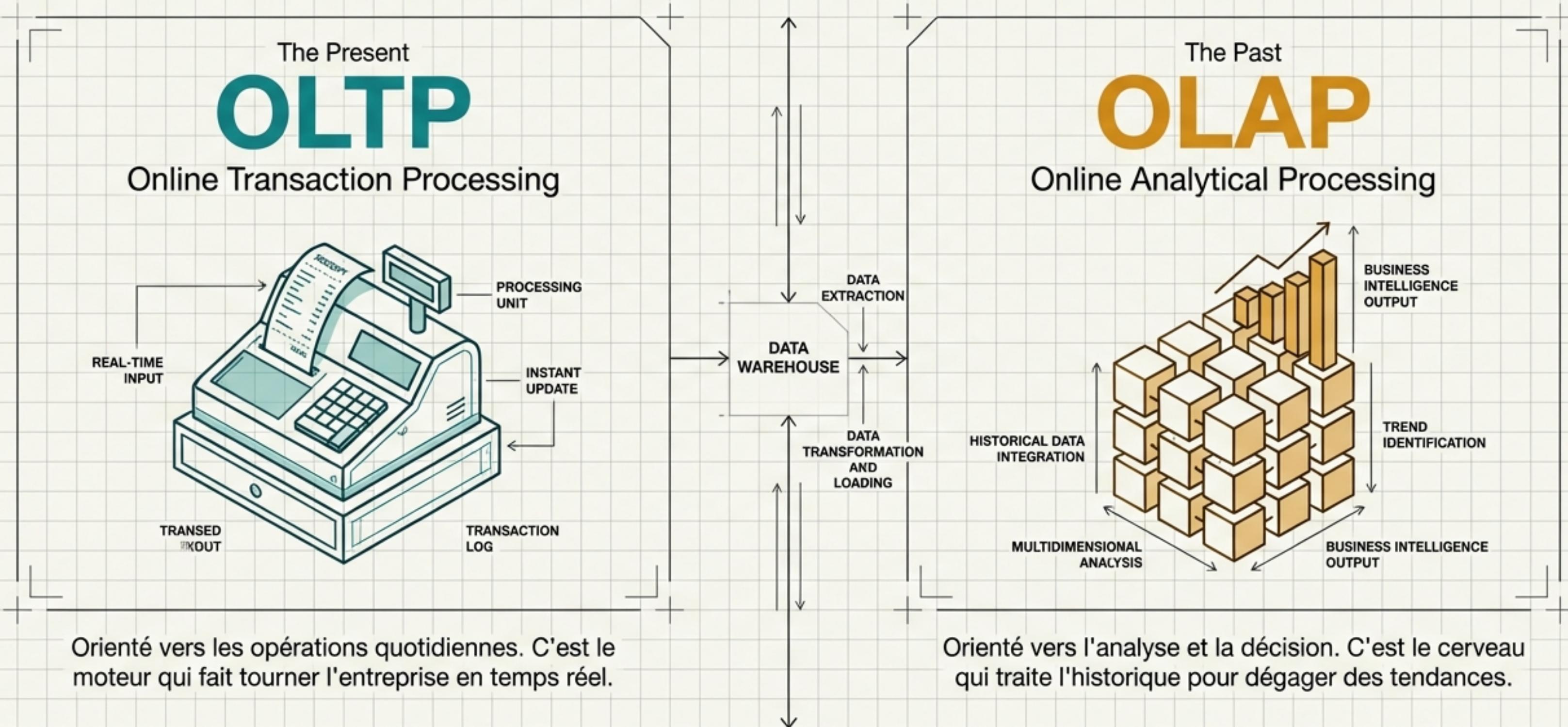
OLTP vs OLAP

Comprendre l'architecture des systèmes de données :
des opérations quotidiennes à l'analyse stratégique.



Deux mondes, deux besoins

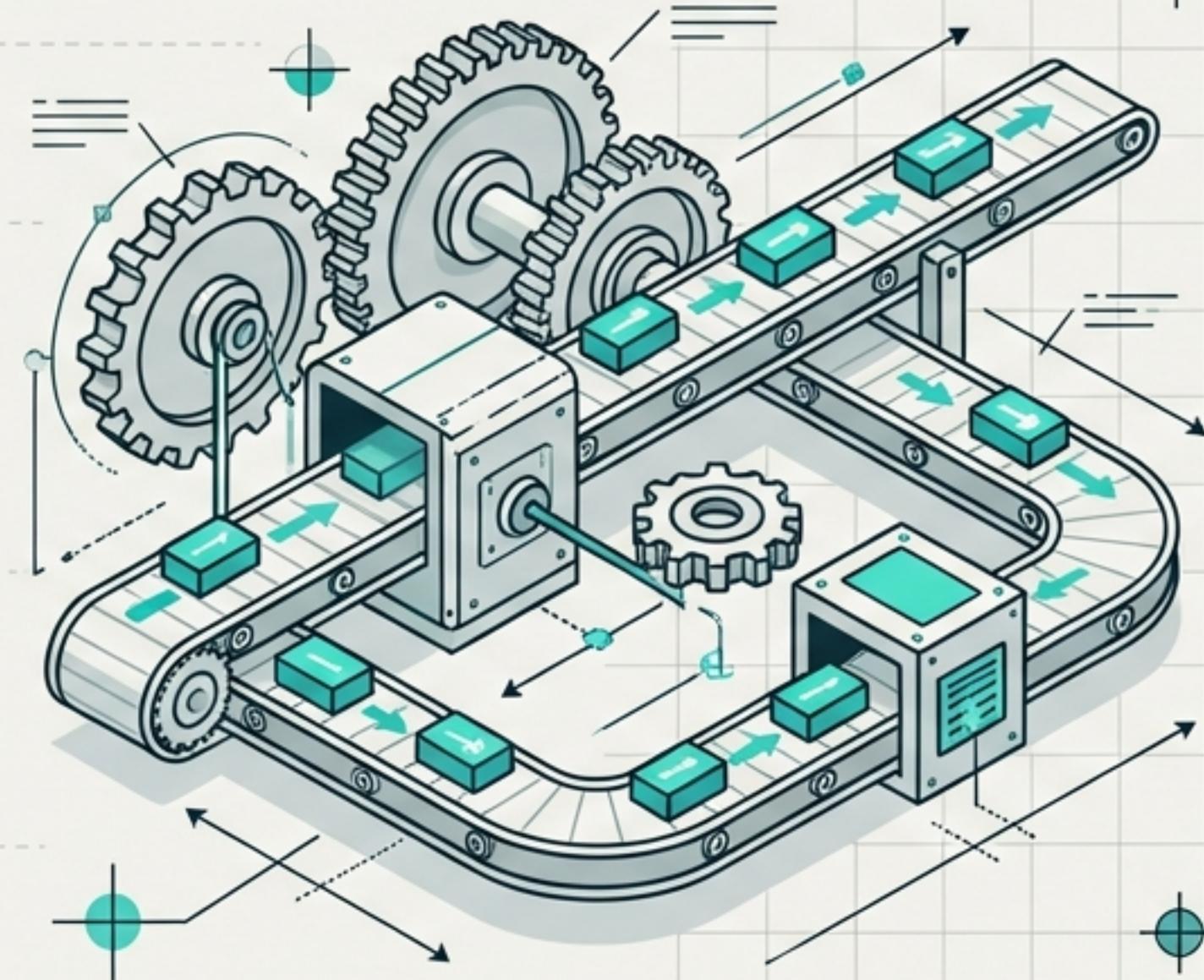
Les systèmes de données ne sont pas universels. Ils se divisent en deux grandes catégories dictées par leur usage final : gérer le présent ou analyser le passé.



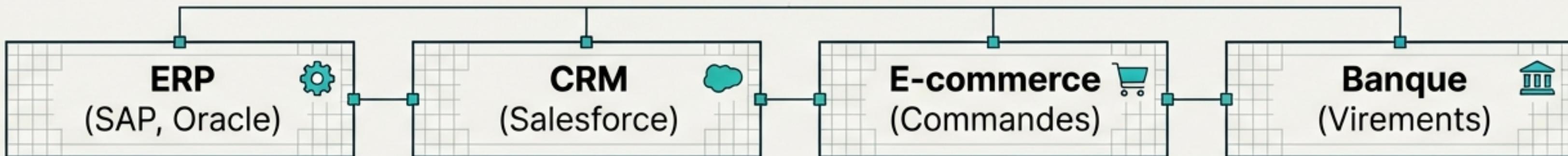
OLTP : Le moteur des opérations

Caractéristiques Clés

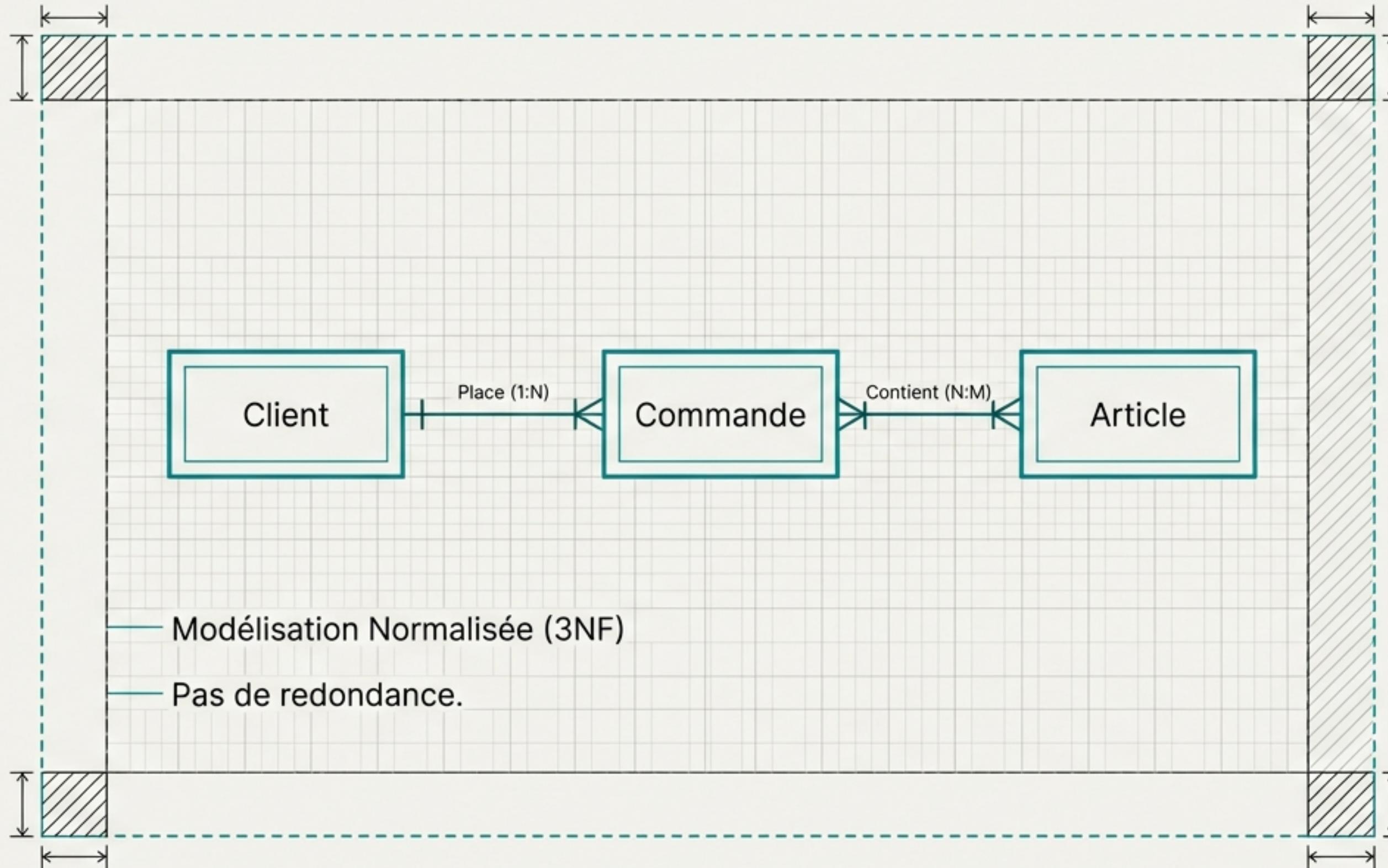
- **Objectif** : Exécuter des tâches atomiques et rapides.
- **Disponibilité** : Critique (99.99%). Si l'OLTP s'arrête, le **business s'arrête**.
- **Utilisateurs** : Employés opérationnels et applications (API, Front-end).



Exemples de systèmes



Architecture OLTP : La précision avant tout



Profil de la Requête

Lecture ou écriture de quelques lignes seulement.

Temps de réponse : Millisecondes.

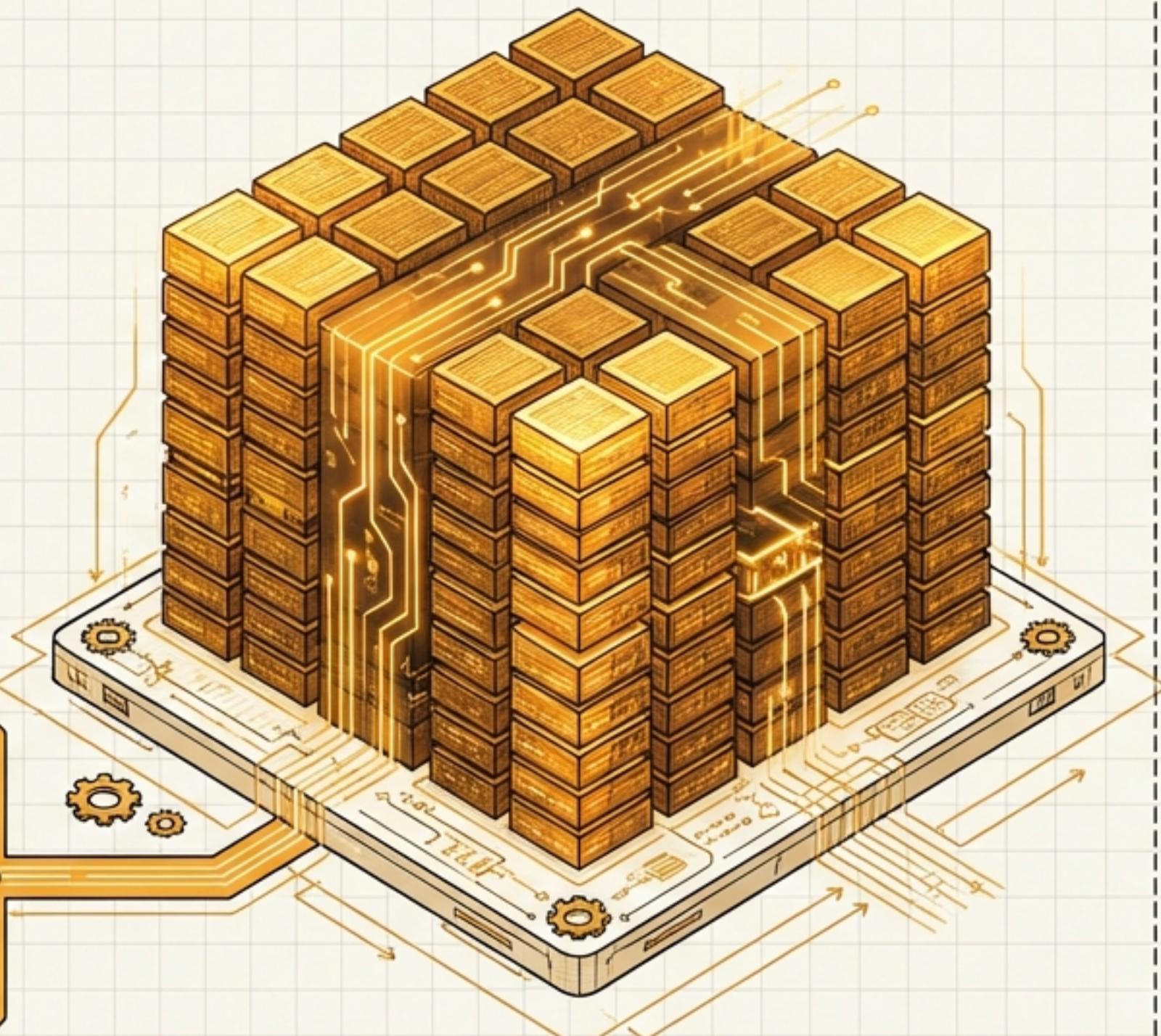
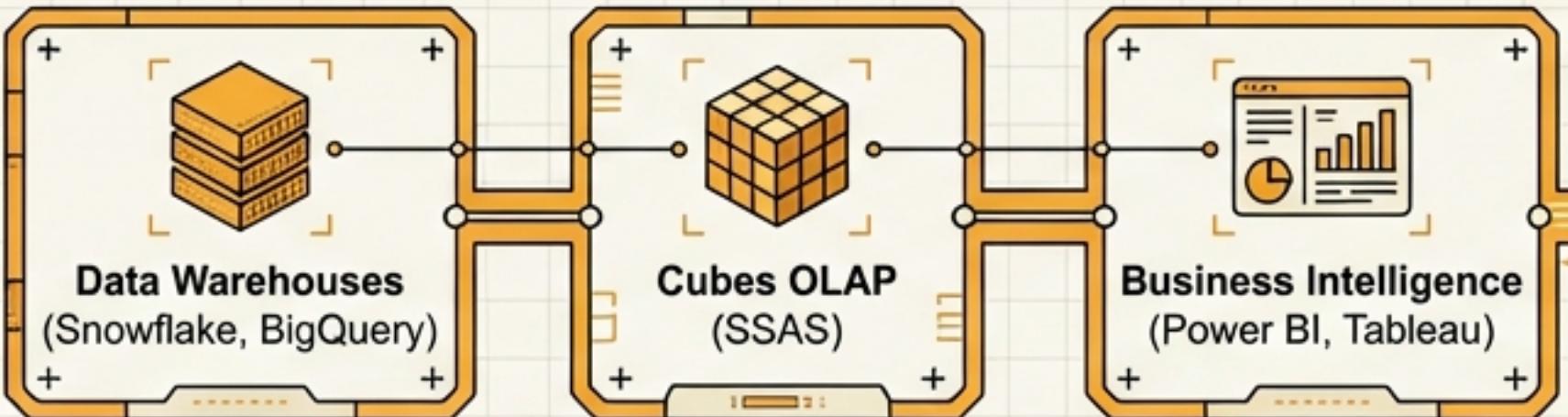
```
SELECT *  
FROM clients  
WHERE id = 101
```

Indexation : B-tree ou Hash pour recherche précise.

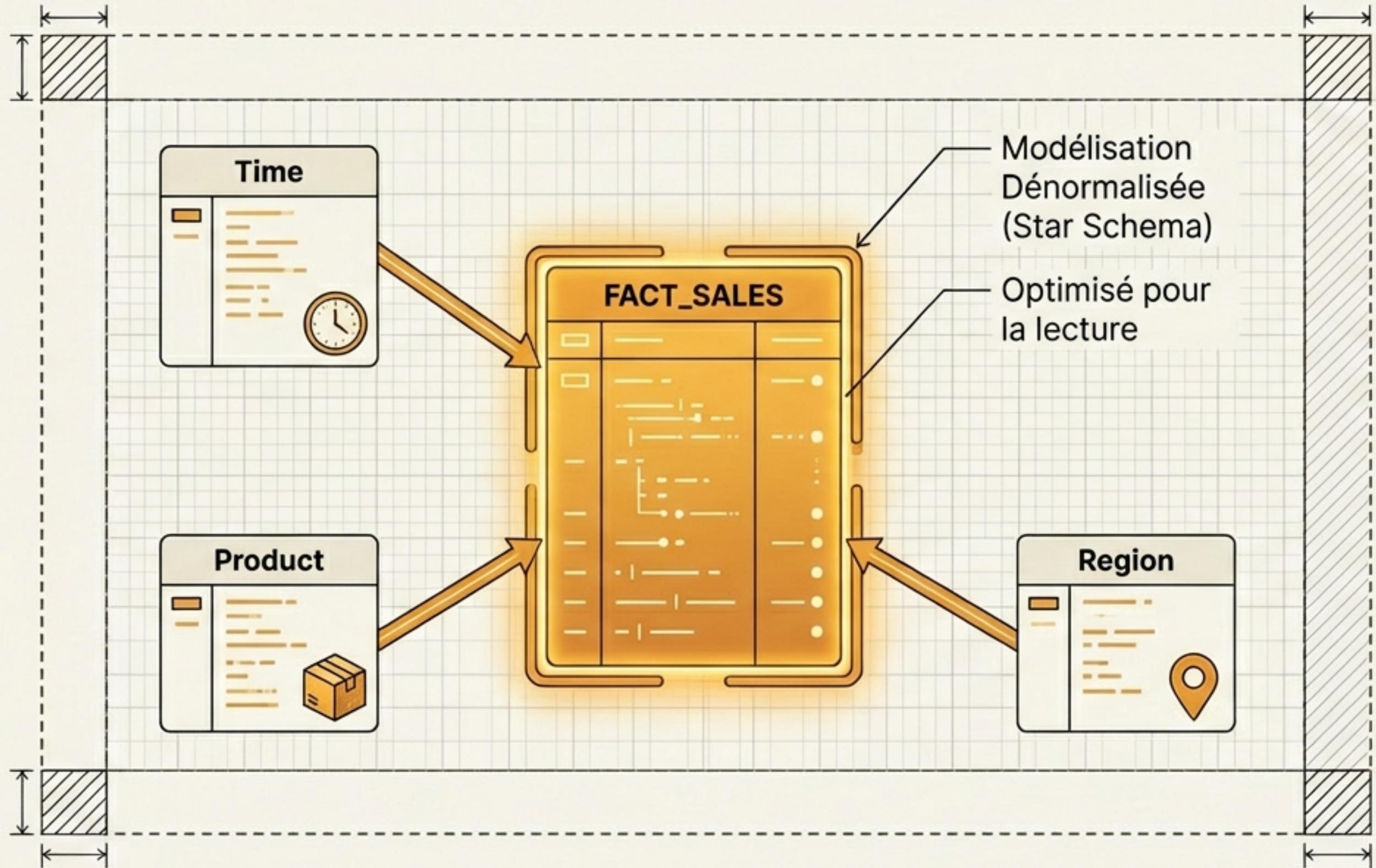
OLAP : Le cerveau analytique

Caractéristiques Clés

- **Objectif** : Aider à la prise de décision stratégique.
- **Données** : Historiques, consolidées et immuables.
- **Utilisateurs** : Data Analysts, Data Scientists, Directeurs.



Architecture OLAP : La puissance de lecture



Profil de la Requête

Agrégations complexes sur des volumes massifs.

Temps de réponse :
Secondes à minutes.

```
+ SELECT region,  
      SUM(ventes)  
   FROM commandes  
 GROUP BY region  
+
```

Stratégie : On sacrifie l'espace disque pour la vitesse de lecture.

Comparaison Détaillée : Le Face-à-Face

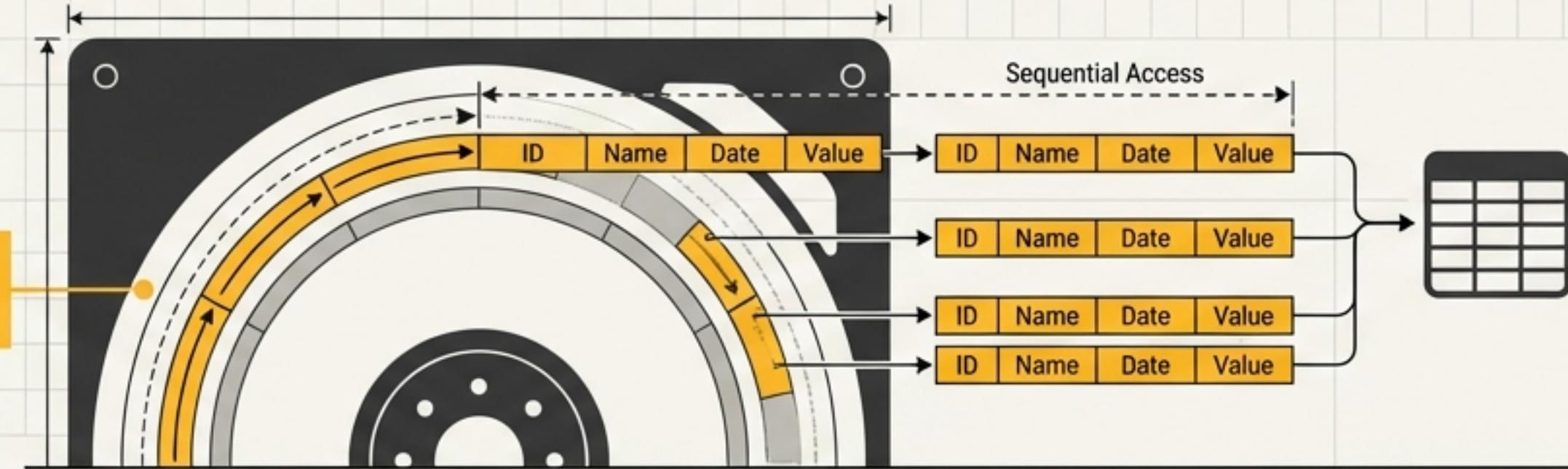
Critère	OLTP	OLAP
Objectif	Opérations (Day-to-day)	Analyse (Reporting)
Volumétrie requête	Faible (Lignes)	Massive (Millions)
Données	Actuelles (Temps réel)	Historiques & Consolidées
Mise à jour	Continue et fréquente	Batch (ETL) périodique
Modélisation	Normalisée (3NF)	Dénormalisée (Star)
Indexation	B-tree	Columnar / Bitmap
Disponibilité	99.99% (Critique)	99.9% (Important)

Sous le capot : Le stockage physique

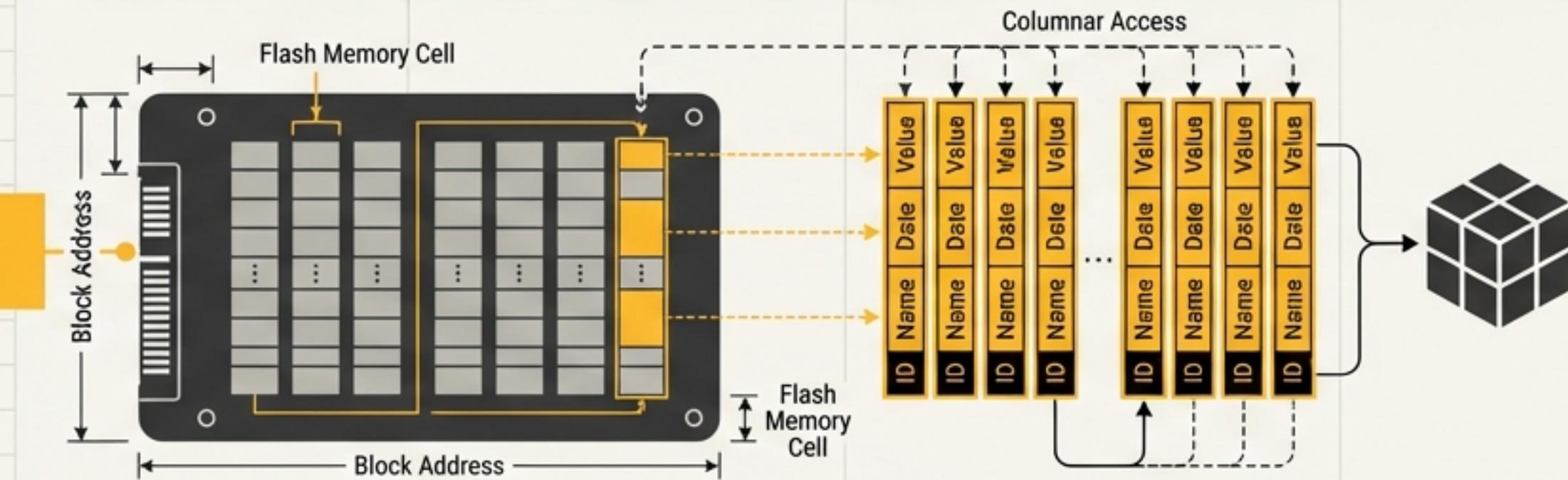
Row vs Column

La différence de performance ne se situe pas seulement dans le logiciel, mais dans la manière physique dont les bits sont écrits sur le disque.

Row-based:
Standard Transactionnel

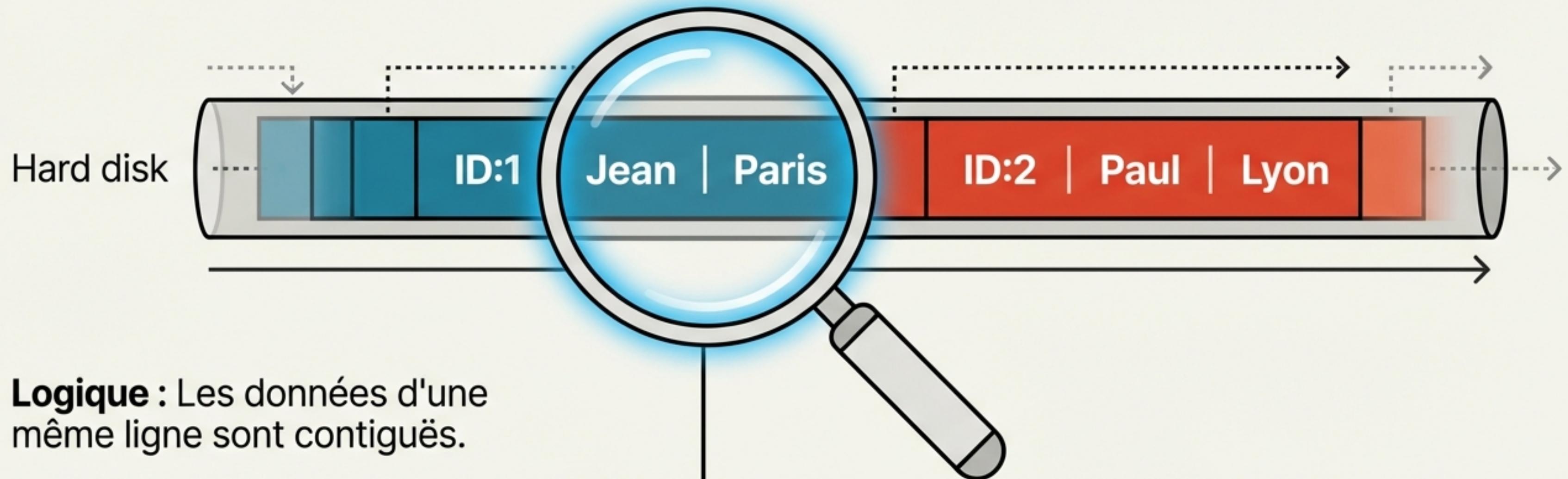


Column-based:
Standard Analytique



Stockage Ligne (Row-Oriented)

Target System: **OLTP** (PostgreSQL, MySQL)



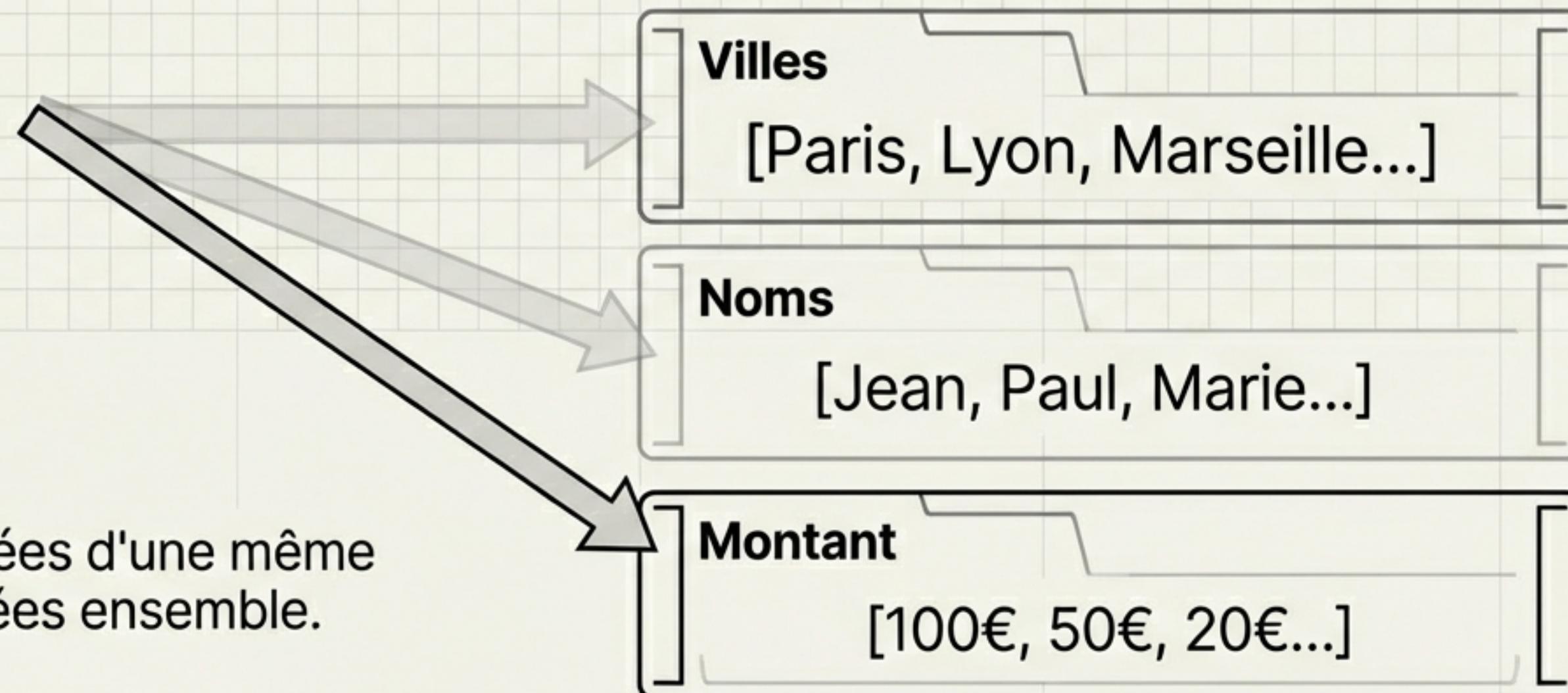
Logique : Les données d'une même ligne sont contiguës.

Avantage : Idéal pour SELECT *.
Récupérer toutes les infos d'un client se fait en une seule lecture.

→ Extraction Rapide :
Lecture Unique

Stockage Colonne (Column-Oriented)

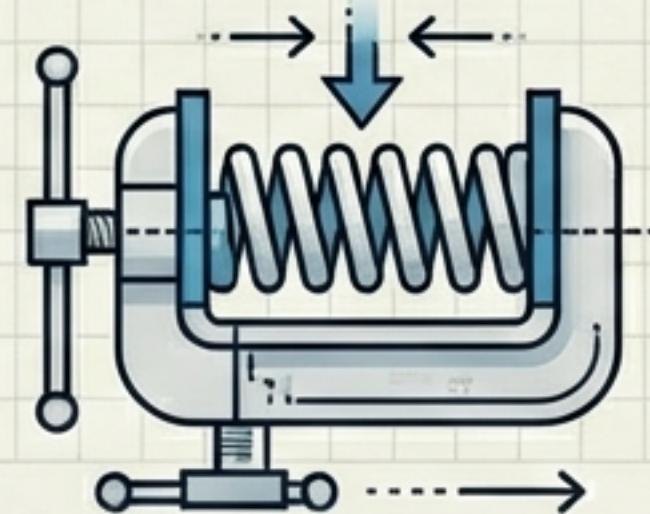
Target System: **OLAP** (Snowflake, BigQuery)



Logique : Les données d'une même colonne sont stockées ensemble.

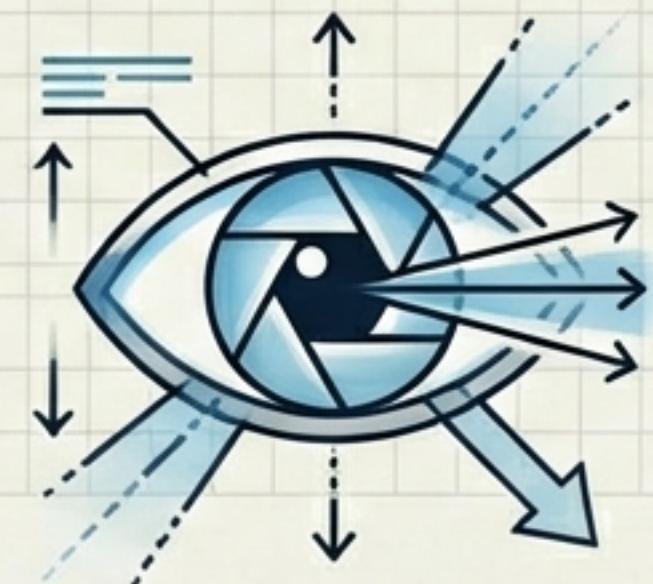
Avantage : Idéal pour SELECT SUM(montant). Le système ne lit que le bloc nécessaire et ignore le reste. Gain I/O immense.

Les super-pouvoirs du stockage colonne



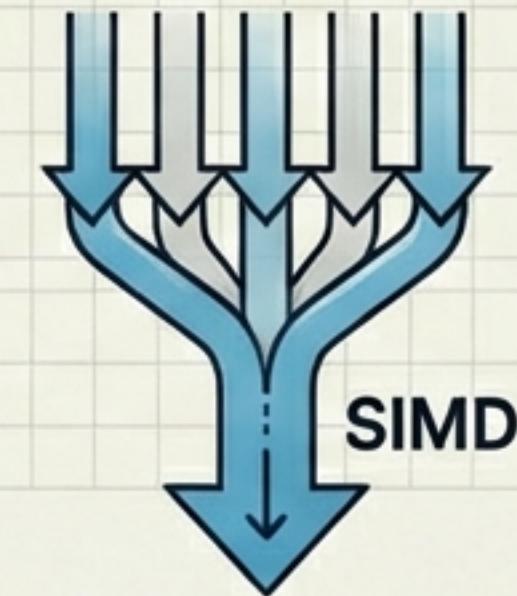
1. Compression Maximale

Des données similaires (ex: 'France', 'France') se suivent. Elles se compressent beaucoup mieux que des données hétérogènes.



2. Lecture Sélective

Le moteur ne charge en mémoire que les colonnes strictement nécessaires à la requête.



3. Vectorisation (SIMD)

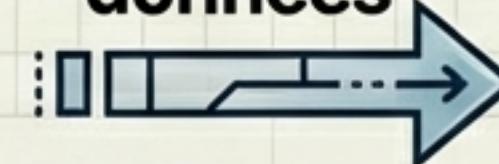
Le processeur applique une opération mathématique sur un bloc entier de valeurs en une seule fois (Single Instruction, Multiple Data).

Cas pratique : Le cycle de vie E-commerce

Midi - La Vente (OLTP)



Extraction de données



Minuit - ETL



Lendemain 9h - Reporting



Midi - La Vente (OLTP)

Validation panier.
Verrouillage stock.
Priorité : Intégrité
immédiate.

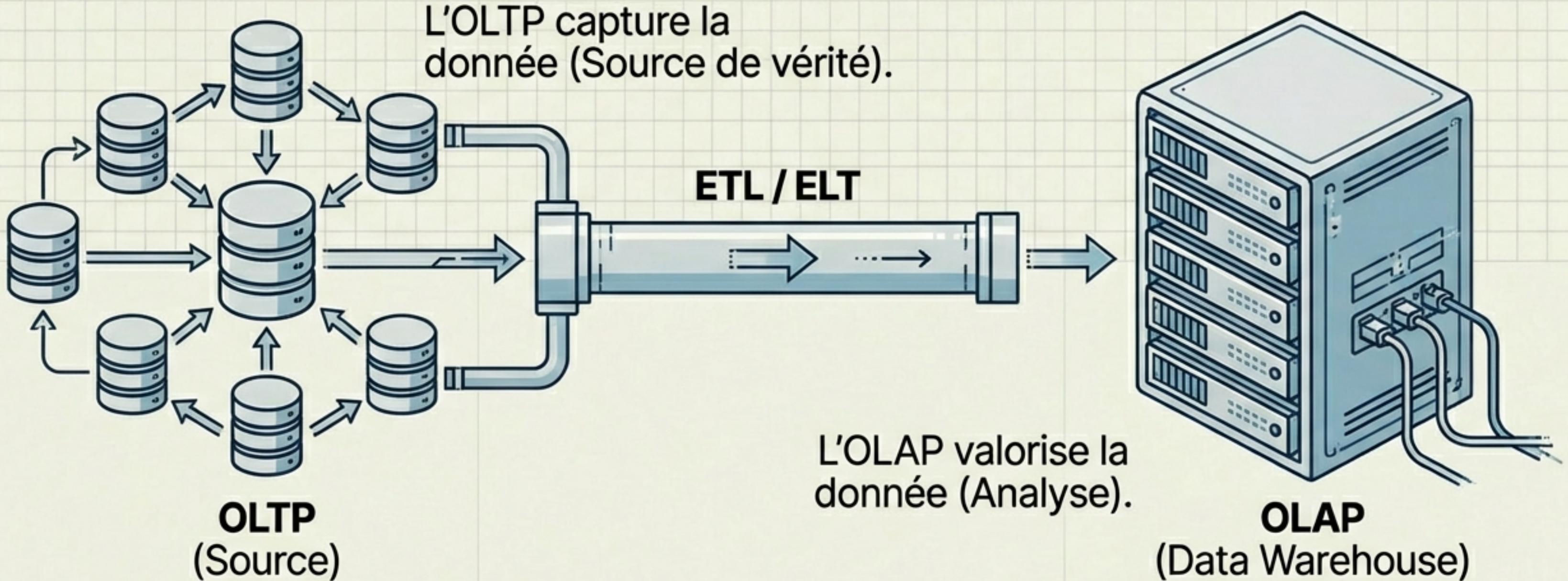
Minuit - ETL

Transformation et
chargement dans
l'entrepôt.

Reporting (OLAP)

Directeur marketing.
Analyse des ventes sur
12 mois.
Priorité : Agrégation
massive.

Un écosystème complémentaire



Règle d'or : Ne jamais lancer de requêtes analytiques lourdes sur la base de production (OLTP).

Ce qu'il faut retenir

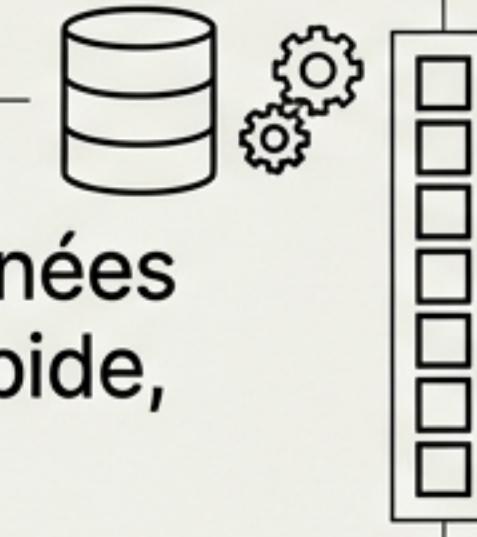
OLTP

Transactions rapides, données fraîches, intégrité, format ligne.



OLAP

Analyses lourdes, données historiques, lecture rapide, format colonne.



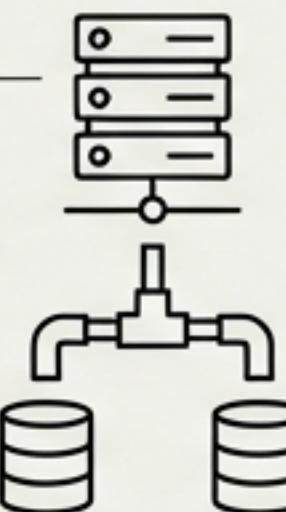
Performance

Le stockage colonne est la clé de la vitesse analytique (agrégations).



Usage

Le Data Warehouse (OLAP) centralise les données produites par l'OLTP.



Conclusion et suite du parcours

Comprendre la distinction OLTP/OLAP est la base de toute architecture Data. Maintenant que nous savons où stocker la donnée pour l'analyse, il faut apprendre à la structurer.

Prochain Module : 03 - Modélisation dimensionnelle