

Modelisation dimensionnelle

Définition

La modélisation dimensionnelle est une technique de conception de bases de données, largement utilisée dans le domaine de l'informatique décisionnelle (ou **Business Intelligence**) pour organiser les données de manière à faciliter l'analyse et l'interrogation des données. Cette approche est principalement utilisée dans les entrepôts de données (**Data Warehouses**) et les systèmes décisionnels pour supporter l'analyse des données historiques et les rapports.

Schéma en étoile vs Schéma en flocon

Schéma en étoile (Star Schema) :

- **Structure** : Dans le schéma en étoile, la **table de faits** (qui contient les mesures ou faits) est au centre et est reliée directement à plusieurs **tables de dimensions** qui apportent le contexte aux données. Ces tables de dimensions ne sont pas normalisées, ce qui signifie que toutes les informations descriptives sont stockées dans une seule table.
- **Caractéristiques** :
 - Simple à comprendre et à interroger (facile pour les utilisateurs finaux).
 - Les tables de dimensions sont **dénormalisées**, donc il peut y avoir une certaine redondance dans les données.
 - Optimisé pour les performances de requêtes, car il nécessite moins de jointures entre les tables.
 - Idéal pour les **requêtes OLAP** simples et rapides.

Exemple : Une table de faits "Ventes" au centre, avec des tables de dimensions comme "Date", "Produit", "Client", et "Région".

Schéma en flocon (Snowflake Schema) :

- **Structure** : Le schéma en flocon est une version normalisée du schéma en étoile. Les tables de dimensions sont divisées en plusieurs sous-tables pour éliminer les redondances. Chaque dimension peut être liée à d'autres tables, créant une structure plus complexe.
- **Caractéristiques** :
 - **Normalisation** des tables de dimensions (les informations sont décomposées en plusieurs tables pour réduire la redondance).
 - Plus complexe, car il nécessite plus de jointures pour interroger les données.
 - Utilise moins d'espace de stockage car la redondance des données est réduite.
 - Moins performant pour les requêtes que le schéma en étoile, en raison des jointures supplémentaires.

***Exemple** : Une table de faits "Ventes" est liée à une table "Produit", qui à son tour est liée à une table "Catégorie de produit", et ainsi de suite.*

Différences principales :

- **Complexité** : Le schéma en étoile est plus simple que le schéma en flocon.
- **Redondance des données** : Le schéma en étoile a plus de redondance, tandis que le schéma en flocon normalise les données pour réduire cette redondance.
- **Performances** : Les requêtes sont généralement plus rapides dans le schéma en étoile car il y a moins de jointures. Le schéma en flocon, bien que plus compact en termes de stockage, peut être plus lent pour les requêtes.

Les types d'OLAP : ROLAP, MOLAP, HOLAP

L'OLAP (Online Analytical Processing) est une technologie permettant de **réaliser des analyses multidimensionnelles** des données. Il existe trois types principaux : **ROLAP**, **MOLAP**, et **HOLAP**.

ROLAP (Relational OLAP) :

- **Structure** : Le ROLAP utilise une base de données **relationnelle** traditionnelle pour stocker les données (comme SQL). Les données sont stockées sous forme de tables dans un entrepôt de données, et les requêtes multidimensionnelles sont traduites en requêtes SQL.
- **Caractéristiques** :
 - Peut gérer de très grands volumes de données (car il s'appuie sur des bases de données relationnelles).
 - Plus **flexible** pour interroger les données complexes (puisque le schéma relationnel est très modulable).
 - Les performances peuvent être plus **lentes** car chaque requête nécessite de nombreuses jointures et agrégations dans une base relationnelle.
 - **Aucune pré-agrégation** des données : les agrégations sont faites au moment de l'interrogation.

Exemple d'outils : SAP Business Objects, MicroStrategy.

MOLAP (Multidimensional OLAP) :

- **Structure** : Le MOLAP utilise des bases de données **multidimensionnelles**, stockant les données sous forme de cubes OLAP. Les données sont pré-agrégées dans ces cubes pour optimiser les performances.
- **Caractéristiques** :
 - Très **rapide** pour l'interrogation, car les données sont **pré-agrégées** et organisées dans un cube multidimensionnel.
 - Utilise plus d'espace de stockage en raison des **agrégations multiples** dans les cubes.
 - Moins flexible car les structures des cubes doivent être définies à l'avance.
 - **Idéal pour les analyses rapides** et répétitives sur des jeux de données relativement constants.

Exemple d'outils : IBM Cognos, Microsoft Analysis Services (SSAS).

HOLAP (Hybrid OLAP) :

- **Structure** : Le HOLAP combine les avantages du ROLAP et du MOLAP. Il stocke les données détaillées dans une base de données relationnelle (comme dans ROLAP) et les agrégations dans des cubes multidimensionnels (comme dans MOLAP).
- **Caractéristiques** :
 - **Meilleure gestion du stockage** : Les données détaillées sont stockées de manière relationnelle (plus d'espace efficace), tandis que les agrégations sont stockées dans des cubes multidimensionnels.
 - Performance **équilibrée** : HOLAP offre des performances rapides grâce aux cubes agrégés tout en permettant l'accès aux données détaillées dans une base relationnelle quand c'est nécessaire.
 - Combinaison de la **flexibilité** du ROLAP et de la **rapidité** du MOLAP.

Exemple d'outils : Microsoft SQL Server Analysis Services (SSAS) en mode hybride.

Différences principales entre ROLAP, MOLAP, et HOLAP

Caractéristique	ROLAP	MOLAP	HOLAP
Stockage des données	Relationnel (tables SQL)	Multidimensionnel (cubes)	Hybride (relationnel + cubes)
Performance	Plus lent (requêtes SQL)	Très rapide (pré-agrégation)	Rapide et flexible
Volume de données	Très grand volume de données	Limité par la taille du cube	Gère de grands volumes
Pré-agrégation	Non	Oui	Partiellement
Complexité de la requête	Flexible, mais complexe	Simple, mais moins flexible	Équilibre entre les deux
Efficacité du stockage	Optimisé (relationnel)	Moins efficace (pré-agrégation)	Équilibre entre stockage et performance

Conclusion

- **Schéma en étoile** et **schéma en flocon** sont des modèles de conception de bases de données qui diffèrent en termes de structure et d'optimisation des performances (étoile pour la simplicité et les performances des requêtes, flocon pour la normalisation et la réduction de la redondance).
- **ROLAP**, **MOLAP**, et **HOLAP** sont des technologies OLAP qui offrent des solutions différentes pour l'analyse multidimensionnelle des données, selon le type de stockage et les besoins en termes de performance et de flexibilité.