

Business Intelligence & Data Warehouse

Résumé du Chapitre 2 : Architecture d'un Système Décisionnel

Professeur : Mme HILAL Imane

Année Universitaire 2025-2026 – 3ème année

1 Architecture de Référence des Systèmes Décisionnels

1.1 Pourquoi créer un nouveau système ?

Les raisons justifiant la création d'un système décisionnel séparé :

1. Réconciliation sémantique

- Dispersion des sources de données au sein d'une entreprise
- Différents codages pour les mêmes données
- L'entrepôt rassemble toutes les informations dans un unique schéma
- Conservation de l'historique des données

2. Performance

- Les données d'aide à la décision nécessitent une autre organisation
- Les requêtes complexes OLAP dégradent les performances des requêtes OLTP

3. Disponibilité

- La séparation augmente la disponibilité
- Une bonne façon d'interroger des sources de données dispersées

4. Qualité des données

- Amélioration et standardisation des données

1.2 Architecture Générale

L'architecture d'un système décisionnel se compose de trois couches principales :

Sources de données → ETL → Entrepôt de données → OLAP → Reporting & Analyses

Composants :

- Sources de données (bases opérationnelles, fichiers Excel, etc.)
- Processus ETL (Extraction, Transformation, Load)
- Data Warehouse et Data Marts
- Metadata Repository
- Outils OLAP
- Outils d'analyse : Reporting, Data Mining

2 Les Sources de Données

2.1 Types de sources

1. Bases de données internes :

- Bases de production de l'entreprise
- Bases créées par les utilisateurs

2. Bases de données externes :

- Données récupérées sur internet (réseaux sociaux, opinions, réclamations en ligne)
- Données de veille concurrentielle du marché
- Données achetées à des fournisseurs de données

3 Le Processus ETL

3.1 Définition

Un **ETL** est un outil d'intégration des données permettant d'automatiser les chargements des données dans le Data Warehouse.

3.2 Fonctionnalités principales

L'ETL permet de :

- Découvrir, analyser et extraire les données à partir de ressources hétérogènes
- Nettoyer et standardiser les données
- Charger les données dans un Data Warehouse

3.3 Les trois phases de l'ETL

3.3.1 1. Extraction (Extract)

- Extraction possible à partir de plusieurs plateformes
- Chargement incrémental ou complet

3.3.2 2. Transformation (Transform)

Les transformations incluent :

- Révisions de format
- Traitement des valeurs NULL
- Valeurs calculées & dérivées
- Fusion des données
- Fractionnement des champs
- Conversion des unités de mesure
- Conversion des dates
- Déduplication

3.3.3 3. Chargement (Load)

- Chargement des tables du Data Warehouse/Data Mart

3.4 Mission principale de l'ETL

1. Problèmes de normalisation :

- Différents encodages, langues
- Différentes abréviations
- Équivalence sémantique
- Différentes normes & unités de mesures

2. Problèmes de données :

- Champs manquants (âge, adresse, etc.)
- Valeurs incorrectes
- Redondance sémantique

3. Incohérences :

- Incohérence des codes
- Incompatibilité référentielle

3.5 Outils ETL du marché

Le marché propose plusieurs outils ETL émergents pour faciliter ces processus d'intégration et de transformation des données.

4 ETL vs ELT

4.1 ELT : Extract, Load, Transform

Caractéristiques :

- Permet de préserver les données à l'état brut
- Offre une plus grande flexibilité d'utilisation des données dans le temps
- Nombreux avantages pour le traitement de grands volumes de données

4.2 Comparaison ETL vs ELT

ETL	ELT
Choix traditionnel	Plus évolutif
Transformation avant chargement	Transformation après chargement
Adapté aux volumes modérés	Adapté aux grands volumes
Données transformées stockées	Données brutes préservées
Moins flexible	Plus flexible dans le temps

5 Le Data Warehouse et les Data Marts

5.1 Définition du Data Warehouse

Un entrepôt de données est une collection de données (une base de données) **intégrée, orientée sujet, historisée et non volatile**, susceptible d'appuyer le processus décisionnel dans une entreprise ou organisation.

5.2 Propriétés d'un Data Warehouse

1. Orienté sujet

- Les données stockées fournissent des informations sur un sujet particulier

2. Intégré

- Intégration de données provenant de multiples sources hétérogènes dans un format cohérent

3. Non volatile

- Les données ne devraient pas changer
- Le but est de permettre d'analyser ce qui s'est passé

4. Historisé

- Les données historiques sont conservées
- Possibilité de récupérer les données de 3, 6, 12 mois, ou même 5, 10 ans

5.3 Les Métadonnées

Définition : Les métadonnées sont les informations relatives à la structure des données, les méthodes d'agrégation et le lien entre les données opérationnelles et celles du Data Warehouse.

Les métadonnées doivent renseigner sur :

- Le modèle de données
- La structure des données telle qu'elle est vue par les développeurs
- La structure des données telle qu'elle est vue par les utilisateurs
- Les sources des données
- Les transformations nécessaires
- Suivi des alimentations

5.4 Les Data Marts

Définition : En français : magasin de données. Les Data Marts sont un sous-ensemble du Data Warehouse où se produit la plupart des activités d'analyse de l'environnement BI.

Caractéristiques :

- Les données sont généralement adaptées pour une capacité ou fonction particulière
- Exemples : analyse de la rentabilité des produits, analyse démographique de la clientèle
- Chaque Data Mart spécifique n'est pas nécessairement valable pour d'autres usages

5.5 Comparaison Data Warehouse vs Data Mart

Data Warehouse	Data Mart
Défini à l'échelle de l'entreprise	Défini à l'échelle départementale
Contient plusieurs domaines	Contient souvent un seul domaine
Contient des informations très détaillées	Peut contenir des données plus résumées
Intègre toutes les sources de données	Intègre les informations d'un sujet donné

6 Approches de Conception

Deux approches principales existent pour la conception d'un entrepôt de données :

6.1 Approche de Ralph Kimball (Bottom-Up)

- Met en avant l'importance des Data Marts
- Le Data Warehouse est une combinaison de différents Data Marts
- Facilite le reporting et l'analyse
- Approche ascendante

6.2 Approche de Bill Inmon (Top-Down)

- Le Data Warehouse est le dépôt centralisé de toutes les données de l'entreprise
- Création d'abord d'un modèle de Data Warehouse normalisé
- Les Data Marts sont ensuite créés sur la base du modèle de l'entrepôt
- Approche descendante

7 Data Warehouse vs Data Lake

7.1 Définition du Data Lake

Un **Data Lake** est une banque de stockage servant à contenir une immense quantité de données brutes dans leur format d'origine jusqu'à ce que l'entreprise en ait besoin.

7.2 Comparaison

Critère	Data Lake	Data Warehouse
Structure des données	Brutes	Traitées/transformées
Objet des données	À déterminer	En service
Utilisateurs	Data scientists	Spécialistes métier
Accessibilité	Accès facile, mises à jour rapides	Modifications plus complexes et coûteuses

Synthèse

Le chapitre 2 présente l'architecture complète d'un système décisionnel, depuis les sources de données jusqu'aux outils de restitution. Le processus ETL joue un rôle central dans la transformation et l'intégration des données. Le Data Warehouse, avec ses propriétés spécifiques (intégré, orienté sujet, historisé, non volatile), constitue le cœur du système décisionnel, tandis que les Data Marts permettent des analyses ciblées par domaine métier.