



**Université d'Alger 1 Benyoucef Benkhedda**  
**Département de Mathématiques et Informatique**



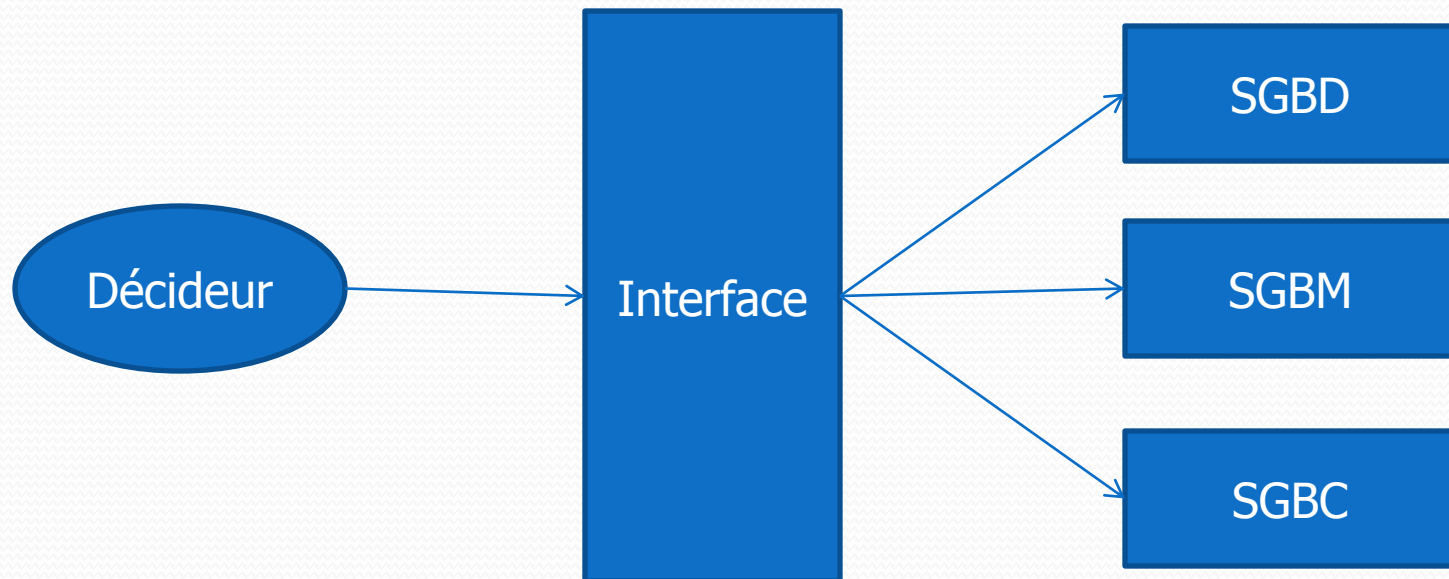
# **Systèmes d'Information d'Aide à la Décision : Les concepts de base**

- **Le niveau:**  
**1ère année master ISII**
- **Année universitaire:**  
**2020/2021**

# Structure d'un système d'aide à la décision

- L'architecture générale du SIAD se compose de cinq parties distinctes :
  - ❑ Le système de gestion de base de données,
  - ❑ le système de gestion de base de modèles,
  - ❑ le système de gestion de base de connaissances (moteur de connaissances),
  - ❑ l'interface utilisateur et les utilisateurs.

# Structure d'un système d'aide à la décision



# L'interface Homme-machine

- C'est un système de gestion et de génération de dialogue entre l'utilisateur et le reste du système
- Est un élément clé des fonctionnalités du système global
- Elle permet au décideur accéder aux données et aux fonctions de calcul,
- Elle permet au système de communiquer le résultat des manipulations.

# Un Système Gestionnaire de Base de Données (SGBD)

- Il stocke, organise, trie et remonte les données pertinentes pour un contexte particulier de décision
- Permettre une recherche rapide de données par des requêtes et des rapports et réaliser des tâches de recherche et de manipulation complexe de données.

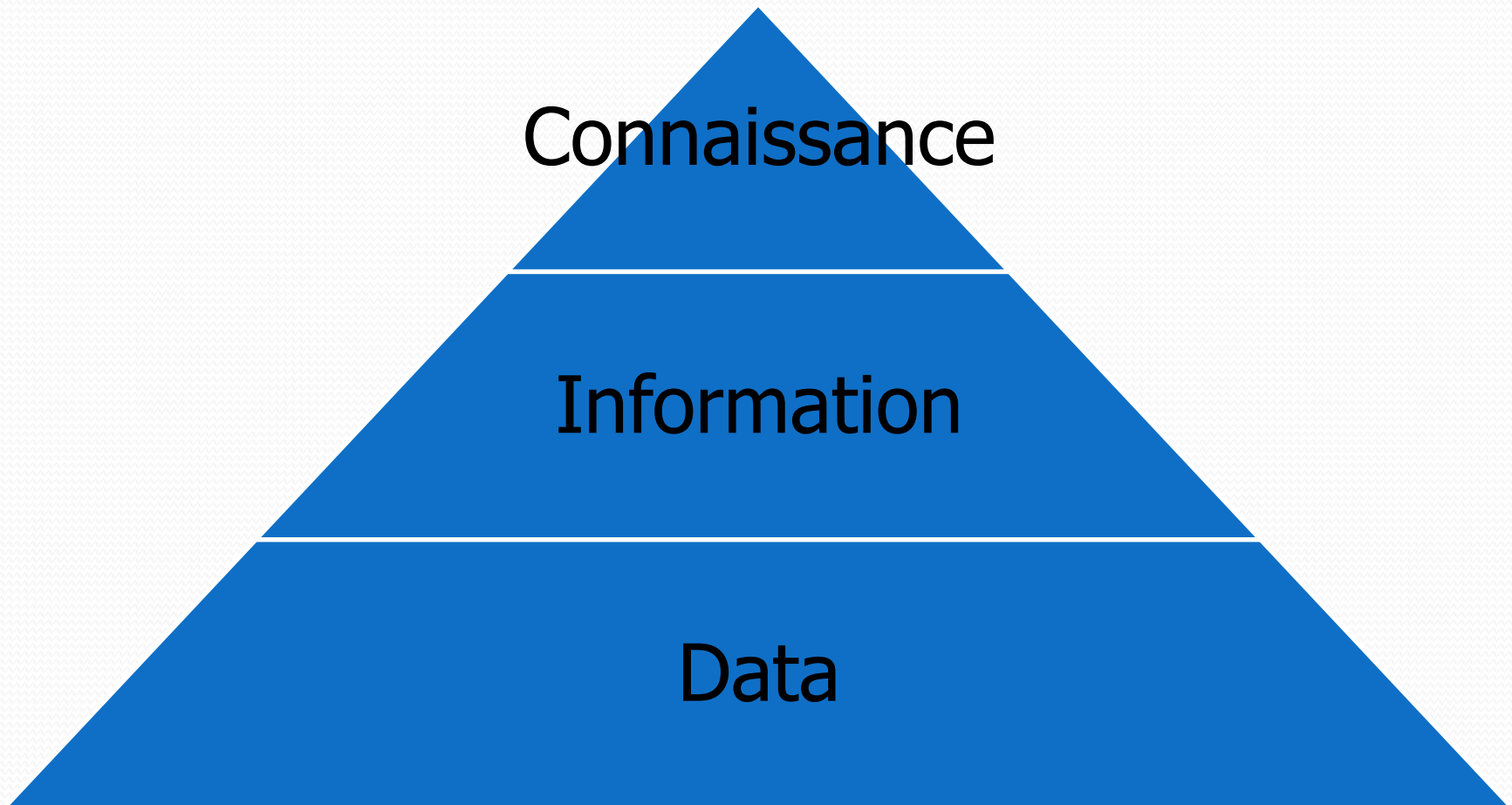
# Un Système Gestionnaire de Base de Modèles (SGBM)

- Il contient l'ensemble des modèles et des procédures de calcul utilisés dans les différents traitements standards des données mis à disposition de l'utilisateur.
- Exemple: outils de programmation mathématiques, outils de prédiction et de simulation, carte géographique

# Un Système de Gestion de Base de Connaissances (SGBC)

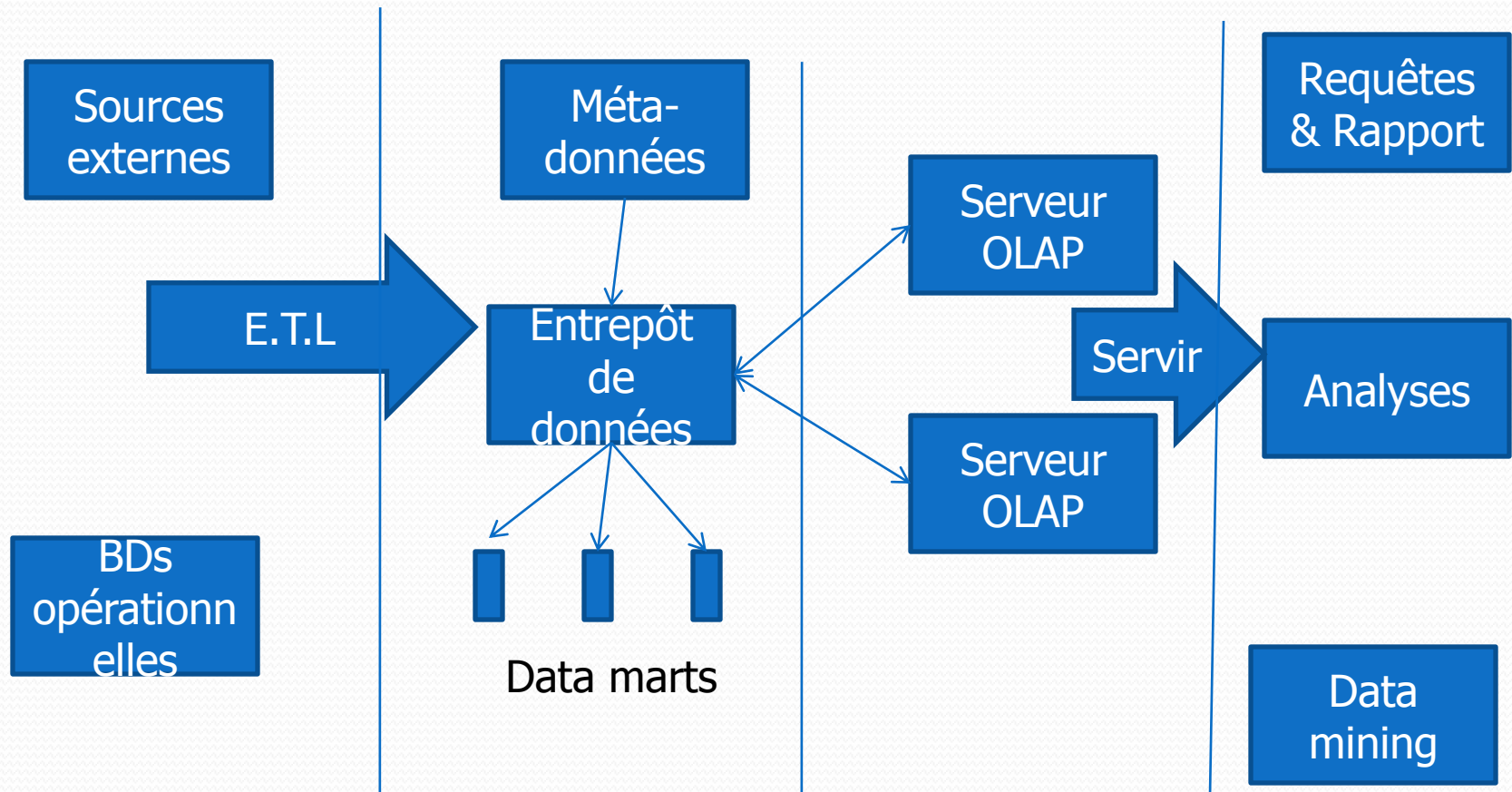
- Les connaissances qui sont nécessaires pour une aide à la décision sont de deux catégories:
  - ❑ Les connaissances de domaine qui servent à construire une représentation du problème
  - ❑ Les connaissances de résolution qui permettent à partir des connaissances de domaine de trouver une solution ou d'apporter des éléments d'aide à la décision.

# La pyramide du système d'aide à la décision





# Architecture du système d'aide à la décision



# Collecte des données

- La collecte (parfois appelée datapumping) est l'ensemble des tâches consistant à détecter, à sélectionner, à extraire et à filtrer les données brutes issues des bases de production

# Collecte des données

- Hétérogénéité des données
- Plusieurs types de sources
- Plusieurs natures d'informations

# Collecte des données

- ETL = Extract, Transform, Load processus:
  - ❑ Extract: Extraire les données de sources hétérogènes
  - ❑ Transform: Transformation des données pour les stocker dans un data warehouse
  - ❑ Load: Chargement des données dans le data warehouse

# Intégration des données

- L'intégration consiste à concentrer les données collectées dans un espace unifié : l'entrepôt de données (Data warehouse)
- Cela inclut:
  - ☐ Le nettoyage et filtrage des données
  - ☐ La validation des données
  - ☐ La synchronisation

# Intégration des données

- Data Mart

Un DataMart (littéralement en anglais magasin de données) est un sous-ensemble d'un DataWarehouse destiné à fournir des données aux utilisateurs, et souvent spécialisé vers un groupe ou un type d'affaire.

# Diffusion des données

- On Line analytical processing (OLAP) est une approche d'analyse multi-dimensionnelle des données.

# But de E.T.L

- L'objectif d'ETL est de produire des données propres, faciles d'accès et qui peuvent être exploitées efficacement par l'analytique.
- La réalisation de l'ETL constitue 70% d'un projet décisionnel en moyenne



# Avantages de E.T.L

- Peut traiter de grandes quantités de données dans une même exécution(traitement en lot);
- Permet des transformations complexes et agrégations sur les données

# Inconvénients de E.T.L

- ❑ Processus de développement long et coûteux
- ❑ Exige de l'espace disque pour effectuer les transformations
- ❑ Unidirectionnel (des sources vers l'entrepôt de données).

# Identification des sources

- Énumérer les items cibles (métriques et attributs de dimension) nécessaires à l'entrepôt de données;
- Pour chaque item cible, trouver la source et l'item correspondant de cette source
- Si plusieurs sources sont trouvées, choisir la plus pertinent
- Si l'item cible exige des données de plusieurs sources, former des règles de consolidation
- Si l'item source referme plusieurs items cibles (ex: un seul champs pour le nom et l'adresse du client), définir des règles de découpage;

# Extraction des données:

- Extraction complète
- ☐ Capture l'ensemble des données à un certain instant (snapshot de l'état opérationnel);
- ☐ Peut être très coûteuse en temps (ex: plusieurs heures/jours).

# Extraction des données:

## Extraction incrémentale

- ❑ Capture uniquement les données qui ont changées ou ont été ajoutées depuis la dernière extraction;
- ❑ Extraction différée ( par lô) Extraction en temps réel

# Transformation des données:

- Types de transformation
- Révision de format
  - Ex: Changer le type ou la longueur de champs individuels
- Décodage de champs
  1. Consolider les données de sources multiples
    - Ex: ['homme','femme'] vs ['M', 'F'] vs [1,2].
  2. Traduire les valeurs cryptiques
    - Ex: 'AC' et 'IN' pour les statuts actif, inactif

# Transformation des données:

- Découpage de champs complexes  
EX. extraire les valeurs prénom et second Prénom à partir d'une seule chaîne de caractères nom Complet
- Fusion de plusieurs champs
- Conversion de jeu de caractères
- Conversion des unités de mesure
- Conversion de dates
- Déduplication

# Chargement des données:

- Chargement initial

Fait une seule fois lors de l'activation de l'entrepôt de données. On charge toutes les données de l'entreprise dans l'entrepôt



# Chargement des données:

- Chargement incrémental
  - Peut être fait en temps-réel ou en lot.
  - Il faudra faire attention dans ce cas à ne charger que les informations nouvelles, et ne pas charger deux fois la même information.

# Chargement des données:

- Considération additionnelles:
  - ❑ Faire les chargements en lot dans une période creuse (entrepôt de données non utilisé);
  - ❑ Considérer la bande passante requise pour le chargement;

# Data Warehouse

- **Définition**

Désigne une BDD utilisée pour collecter, ordonner, journaliser et stocker des informations provenant de base de données opérationnelles et fournir ainsi un socle à l'aide à la décision en entreprise.

- Collecter: Récupérer l'information(ETL)
- Ordonner: Structurer l'information dans le but de la prise de décision
- Journaliser: Stocker l'historique des données

# Avantages

- Permettre l'exécution de requêtes longues, bloquantes, sur des données opérationnelles
- Intégrer des données provenant de multiples systèmes sources
- Améliorer la qualité des données
- Fournir un seul modèle de données commun pour toutes les données d'intérêt, indépendamment de la source de données

# Etapes de création

- Interview utilisateurs ( les besoins)
  - Une phase préalable d'interviews auprès des utilisateurs est nécessaire
- Modélisation de DW
  - Organiser les données en utilisant un des modèles de DW
- Modélisation de Data Mart
  - Construction des données qui vont être utilisées au sein dusevice

# OLAP

- Ensemble des technologies qui, se basant sur une représentation multi-dimensionnelle des données, permet aux analystes et décideurs de:
- Traiter leurs données de façon analytique, interactive (sessions) et rapide.
- Voir les données de l'entreprise sous plusieurs angles (dimensions)

# OLAP

- Les systèmes OLAP permettent non seulement de répondre aux questions de type "qui" et "quoi", mais permettent aussi de répondre aux "que se passe t il si, et aux "pourquoi?"

# OLAP

- OLAP permet la prise de décision sur des actions futures
- OLAP et DW sont complémentaires. Un DW stocke et gère les données. OLAP transforme les données de l'entrepôt en informations stratégiques



# OLAP

- Des données (liées à l'activité de l'entreprise) accessibles.
- Des données compréhensibles
- Des données agrégées (résumés d'informations).
- Un mode de requêtage simple.

# Inconvénients

- Des applications volumineuses.
- Des temps de réponse parfois longs