

**ESI**

École des Sciences de l'Information

Rabat, Maroc

# Correction de l'Examen Final Data Warehouse / Business Intelligence

Filières : ISITD et IIN / ICSD

**Année Universitaire :** 2024-2025

**Professeur :** Mme HILAL Imane

**Durée :** 1 heure

**Barème :** /20

**Correction Détailée**  
avec explications et schémas

## Contents

<b>1 Partie I : QCM (10 Points)</b>	<b>2</b>
<b>2 Partie II : Conception d'un Data Warehouse (10 Points)</b>	<b>8</b>
2.1 Question 1 : Architecture du Système Décisionnel . . . . .	8
2.2 Question 2 : Schéma en Étoile du Data Warehouse . . . . .	10
2.3 Question 3 : Schéma en Flocon de Neige . . . . .	12

## 1 Partie I : QCM (10 Points)

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
a	a	a	b	b	b	a	a	a	b

### Question 1

La Business Intelligence inclut l'utilisation des techniques Data Mining pour extraire le maximum de pattern et de connaissance des informations décentralisées de l'entreprise.

- a. Vrai
- b. Faux

### Réponse et Justification

#### Réponse : a. Vrai

La Business Intelligence (BI) englobe l'ensemble des solutions informatiques d'aide à la décision. L'architecture de référence d'un système décisionnel comprend effectivement le **Data Mining** comme composant d'analyse avancée, aux côtés de l'OLAP et du Reporting. Le Data Mining permet d'extraire des patterns et des connaissances cachées à partir des données consolidées dans l'entrepôt de données.

### Question 2

L'ETL peut être utilisé comme outils de Data Integration en dehors de l'architecture BI.

- a. Vrai
- b. Faux

### Réponse et Justification

#### Réponse : a. Vrai

L'ETL (Extract, Transform, Load) est un **outil d'intégration de données** générique qui peut être utilisé dans divers contextes :

- Migration de données entre systèmes
- Synchronisation de bases de données
- Consolidation de données pour des applications non-BI
- Intégration de systèmes hétérogènes

Bien que l'ETL soit central dans l'architecture BI, son utilisation dépasse ce seul cadre.

### Question 3

L'ELT est réservé principalement dans les cas des données volumineuses et se déploie généralement en cloud.

- a. Vrai  
 b. Faux

### Réponse et Justification

#### Réponse : a. Vrai

L'ELT (Extract, Load, Transform) se distingue de l'ETL par l'ordre des opérations :

- **ETL** : Transformation avant chargement (approche traditionnelle)
- **ELT** : Chargement des données brutes puis transformation dans le système cible

L'ELT est particulièrement adapté aux environnements cloud et Big Data car :

- Il préserve les données brutes pour une flexibilité d'utilisation future
- Il exploite la puissance de calcul des plateformes cloud (Snowflake, BigQuery, etc.)
- Il est plus évolutif pour les grands volumes de données

#### Question 4

Le chargement complet du data warehouse permet d'alimenter les derniers changements effectués au niveau des sources dans l'entrepôt de données.

- a. Vrai
- b. Faux

#### Réponse et Justification

##### Réponse : b. Faux

Il faut distinguer deux types de chargement :

- **Chargement complet (Full Load)** : Recharge *toutes* les données depuis les sources, écrasant les données existantes
- **Chargement incrémental (Incremental Load)** : Charge uniquement les *changements* (nouvelles données, modifications, suppressions)

La phrase décrit le comportement du **chargement incrémental**, pas du chargement complet. Le chargement complet est coûteux en temps et ressources car il retraite l'intégralité des données.

#### Question 5

La fréquence du chargement du datawarehouse définit la granularité des indicateurs de la table de fait.

- a. Vrai
- b. Faux

#### Réponse et Justification

##### Réponse : b. Faux

Ce sont deux concepts distincts :

- **Fréquence de chargement** : Rythme auquel les données sont rafraîchies (quotidien, hebdomadaire, temps réel...)
- **Granularité** : Niveau de détail des mesures dans la table de faits (transaction, jour, mois...)

La granularité est définie lors de la **conception** du datawarehouse et détermine le niveau de détail accessible pour l'analyse. Par exemple, une granularité à la transaction permet plus de détail qu'une granularité mensuelle. La fréquence de chargement n'affecte pas ce niveau de détail.

**Question 6**

L'utilisation des outils OLAP tel que les ROLAP et MOLAP est obligatoire dans l'architecture Data Warehouse avant la restitution des données sous format tableaux de bord.

- a. Vrai
- b. Faux

**Réponse et Justification****Réponse : b. Faux**

L'utilisation d'OLAP n'est **pas obligatoire**. L'architecture décisionnelle peut inclure :

- Des tableaux de bord connectés directement au datawarehouse via SQL
- Des outils de reporting sans couche OLAP intermédiaire
- Des solutions de visualisation modernes (Power BI, Tableau) qui peuvent interroger directement les données

OLAP (ROLAP, MOLAP, HOLAP) est une **option** pour optimiser les analyses multidimensionnelles complexes, mais pas une obligation architecturale.

**Question 7**

Pour appliquer le Slowly Changing Dimension, la clé significative peut être dupliquée dans la table de dimension.

- a. Vrai
- b. Faux

### Réponse et Justification

**Réponse : a. Vrai**

Dans le **SCD Type 2** (Slowly Changing Dimension), lorsqu'un attribut change :

- Une **nouvelle ligne** est créée avec une nouvelle clé de substitution (clé non significative)
- La **clé significative** (business key, provenant du système source) reste identique et est donc dupliquée
- Cela permet de conserver l'historique complet des changements

Clé_surr	ID_client	Nom	Ville	Valide
1	C001	Dupont	Paris	Non
2	C001	Dupont	Lyon	Oui

Ici, ID\_client (clé significative) est dupliquée car le client a déménagé.

### Question 8

Pour mieux répondre aux requêtes du décideur, on peut regrouper un ensemble d'opérateur OLAP à la fois dans une seule requête.

- a. Vrai  
 b. Faux

### Réponse et Justification

**Réponse : a. Vrai**

Les opérations OLAP peuvent être combinées dans une même requête pour des analyses sophistiquées :

- **Drill-down** : Descendre dans la hiérarchie (Année → Mois → Jour)
- **Roll-up** : Remonter dans la hiérarchie (agrégation)
- **Slice** : Sélectionner une tranche du cube
- **Dice** : Sélectionner un sous-cube
- **Pivot** : Rotation des axes d'analyse

Une requête complexe peut combiner un slice sur une période, un drill-down sur les produits, et un roll-up sur les régions.

### Question 9

Le langage MDX permet d'appliquer les opérations sur OLAP pour enrichir les rapports des décideurs.

- a. Vrai
- b. Faux

### Réponse et Justification

**Réponse : a. Vrai**

**MDX (Multidimensional Expressions)** est le langage de requête standard pour interroger les cubes OLAP :

- Conçu spécifiquement pour les structures multidimensionnelles
- Permet d'effectuer des calculs complexes sur les dimensions et mesures
- Utilisé par Microsoft Analysis Services, Oracle Essbase, etc.
- Syntaxe adaptée aux opérations OLAP (members, sets, tuples)

MDX enrichit les capacités analytiques au-delà de ce que SQL peut offrir sur des données relationnelles.

### Question 10

La construction des datamarts par département et ensuite les englober dans un seul datawarehouse est une approche coûteuse en termes de temps et d'efforts.

- a. Vrai
- b. Faux

### Réponse et Justification

**Réponse : b. Faux**

Cette description correspond à l'approche **Kimball (Bottom-Up)** qui est justement considérée comme **moins coûteuse** en temps initial :

Critère	Kimball (Bottom-Up)	Inmon (Top-Down)
Temps initial	Court	Long
Coût initial	Faible	Élevé
Complexité	Modérée	Élevée
ROI rapide	Oui	Non

L'approche **Inmon (Top-Down)**, qui construit d'abord un datawarehouse centralisé puis les datamarts, est celle qui est plus coûteuse en temps et efforts initiaux.

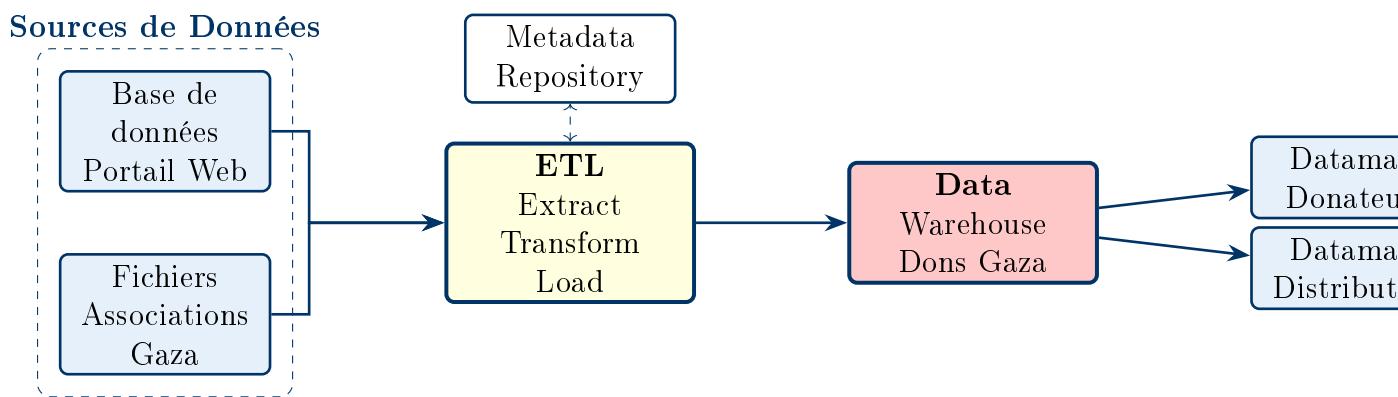
## 2 Partie II : Conception d'un Data Warehouse (10 Points)

**Contexte :** Portail web pour la collecte d'aides financières destinées aux victimes à Gaza. Le tableau de bord doit afficher :

- Le total du montant des dons collectés
- Par donateur (âge, sexe, pays, région)
- Par date
- Par type et région d'aide

**Sources :** Base de données du portail + Fichiers des associations à Gaza

### 2.1 Question 1 : Architecture du Système Décisionnel



**Description des couches :**

#### 1. Sources de Données

- *Base de données du portail web* : Contient les informations des donateurs, montants, dates de dons
- *Fichiers des associations* : Fichiers Excel/CSV avec les données de distribution des aides sur place

#### 2. Couche ETL (Extract, Transform, Load)

- *Extraction* : Récupération des données depuis les deux sources hétérogènes
- *Transformation* : Nettoyage, standardisation des formats, conversion des devises, déduplication, traitement des valeurs NULL
- *Chargement* : Alimentation du Data Warehouse selon la granularité définie

#### 3. Metadata Repository

- Stocke les métadonnées : modèle de données, sources, transformations appliquées, suivi des alimentations

#### 4. Data Warehouse

- Entrepôt centralisé, **orienté sujet** (Dons Gaza), **intégré**, **historisé** et **non volatile**
- Contient la table de faits et les tables de dimensions

#### 5. Datamarts

- *Datamart Donateurs* : Vue analytique sur les profils des donateurs
- *Datamart Distribution* : Vue analytique sur la distribution des aides

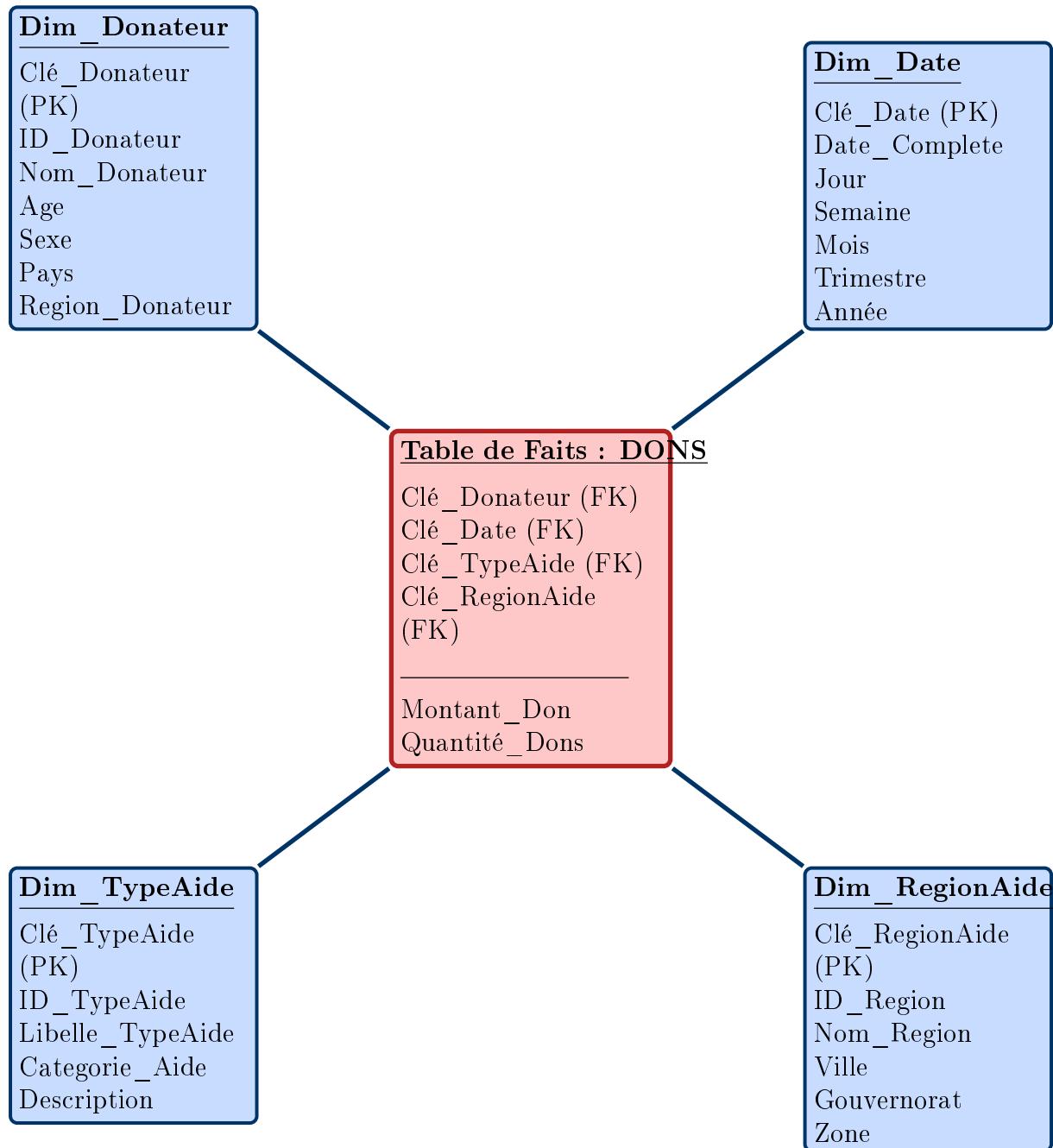
#### 6. Couche OLAP

- Cubes multidimensionnels pour analyses interactives (drill-down, roll-up, slice, dice)
- Langage MDX pour requêtes complexes

#### 7. Couche Restitution

- *Tableaux de bord* : Visualisation interactive du suivi des dons et distributions
- *Rapports* : Documents de synthèse périodiques pour les décideurs

## 2.2 Question 2 : Schéma en Étoile du Data Warehouse



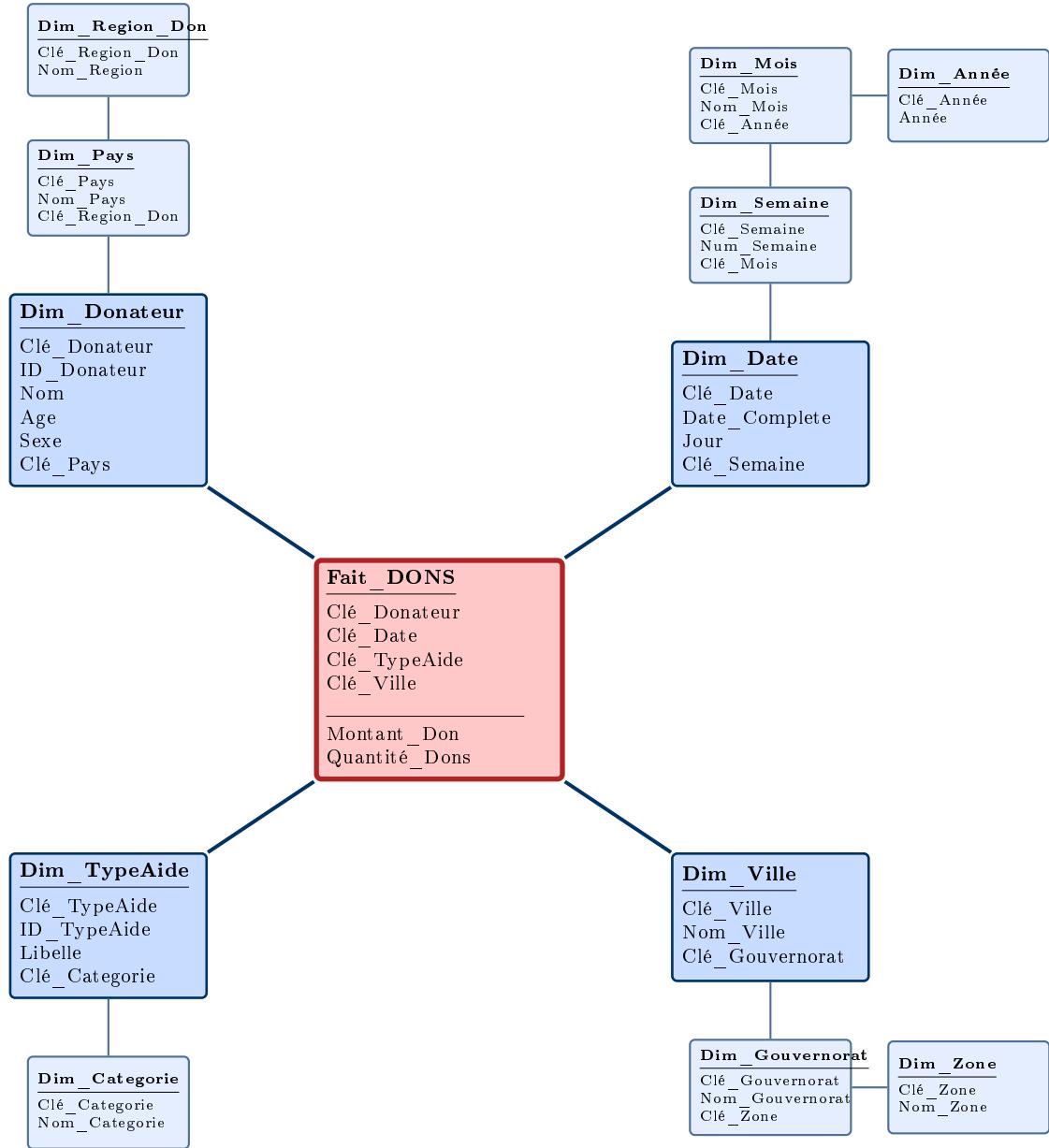
### Justification du modèle :

- **Table de Faits (DONS)** : Granularité au niveau de chaque don individuel
  - Mesures : Montant\_Don (somme des donations), Quantité\_Dons (nombre de dons)
  - Clés étrangères vers les 4 dimensions d'analyse
- **Dim\_Donateur** : Répond au besoin d'analyse par profil donneur (âge, sexe, pays, région)

- **Dim \_ Date** : Permet l'analyse temporelle avec hiérarchie complète
- **Dim \_ TypeAide** : Classification des types d'aide (alimentaire, médicale, financière...)
- **Dim \_ RegionAide** : Zones géographiques de distribution à Gaza

## 2.3 Question 3 : Schéma en Flocon de Neige

Le schéma en flocon normalise les dimensions pour éliminer les redondances :



Transformations appliquées (Normalisation) :

1. **Dim\_Donateur** → Normalisation géographique :
  - Extraction de **Dim\_Pays** et **Dim\_Region\_Donateur**
  - Hiérarchie : Région → Pays → Donateur
2. **Dim\_Date** → Normalisation temporelle :
  - Extraction de **Dim\_Semaine**, **Dim\_Mois**, **Dim\_Annee**
  - Hiérarchie : Année → Mois → Semaine → Date
3. **Dim\_TypeAide** → Normalisation catégorielle :

- Extraction de Dim\_Categorie
- Hiérarchie : Catégorie → Type d'Aide

#### 4. Dim\_RegionAide → Normalisation géographique Gaza :

- Extraction de Dim\_Ville, Dim\_Gouvernorat, Dim\_Zone
- Hiérarchie : Zone → Gouvernorat → Ville

**Avantages du Flocon :**

- Réduction de la redondance des données
- Économie d'espace de stockage
- Maintenance facilitée des référentiels

**Inconvénients :**

- Requêtes plus complexes (plus de jointures)
- Performance potentiellement impactée