

BLOC : BIG DATA

WORKSHOP : MODELISATION MULTI- DIMENSIONNELLE

Introduction :

Suite aux activités faites lors des séquences de préparation du bloc “ Big Data”, vous avez été rappelé aux notions liées aux bases de données relationnelles qui sont liées aux systèmes OLTP (traitement transactionnel en ligne) pour les bases de données opérationnelles/transactionnelles.

Afin de pouvoir modéliser un datawarehouse/datamart (entrepôt de magasin de données), vous êtes amenés à utiliser les systèmes OLAP (traitement analytique en ligne) et à modéliser des données dimensionnelles/de rapport en manipulant des tables de faits et des tables de dimensions selon différents contextes. Un passage des différents modèles de données est important également pour une bonne implémentation de données.

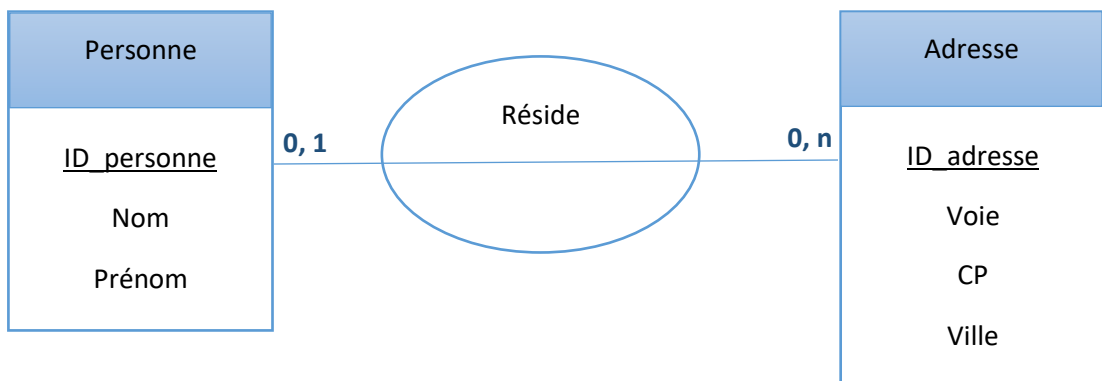
Important :

Il est à noter que les tuteurs doivent accompagner les élèves dans la compréhension de chaque concept lié à la modélisation décisionnelle. Cette dernière étant nouvelle, il est fortement conseiller de réaliser un résumé de ces concepts avant d’entamer le ws.

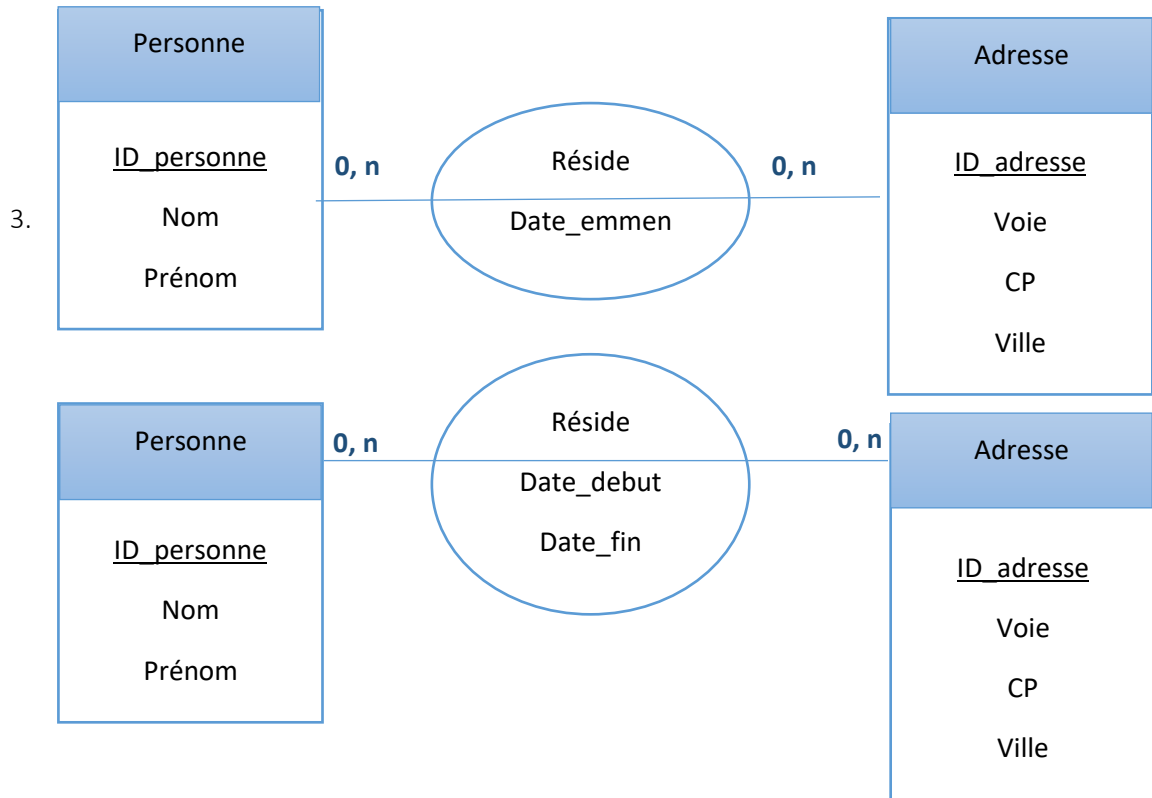
Exercice 1 :

Dans le but de maîtriser le passage d’un modèle conceptuel (ou modèle entité-association) à un modèle logique de données, on demande de conclure les modèles logiques de données à partir des modèles conceptuels suivants :

1. Modèle 1 :



2. Modèle 2 :



Exercice 2 :

Considérons le contexte suivant :

« Soit ERAX une entreprise industrielle qui stocke des produits dans différents points de stockage. Lors de l'achat d'un ou plusieurs produits, les employés d'ERAX font enregistrer les clients afin de pouvoir bénéficier des promotions dans le temps. »

1. A l'aide des notions de "Table de fait" (les données qui peuvent être mesurées) et de "Table de dimension" (ce qui est stocké), présenter une modélisation dimensionnelle en étoile qui aidera à mesurer les données.¹
2. Apportez les changements nécessaires au modèle déjà proposé afin de le rendre modèle en flocon.
3. Citer les principales avantages et inconvénients des deux modèles.²

¹ Une clé primaire identifie de manière unique chaque occurrence de la table, alors qu'une clé étrangère sert à créer des liens entre les tables.

² A explorer en indépendance.

Indication : nous aimerions analyser dans le temps, les quantités achetées ainsi que les coûts associés pour chaque produit lié au client qui l'a acheté et dans chacun des magasins avec la promotion associée.

Données

Nom de la promotion, Type du prix, Nom magasin, Adresse_magasin, Libellé_produit, Marque Groupe_produit, Nom_client, Prénom, Adresse_client, Profile_client, Pays, Nom région, nom_employé, mail_employé....

Exercice 3 :

Considérons le contexte suivant :

« Durant les vacances d'été, Thierry voyagera à l'étranger. Par l'occasion, il souhaite réserver une chambre dans un hôtel. Il a fait une recherche et il s'est rendu compte que les options envisagées caractérisent les hôtels qui ont un bon classement en terme d'étoiles. Ces hôtels font généralement partie d'une chaîne hôtelière. Certes, les caractéristiques recherchées peuvent être trouvés dans plusieurs hôtels. Toutefois, il n'y a forcément pas de disponibilités de chambres. Par ailleurs, le tarif journalier du type de chambres envisagées est au-delà de la compétence financière de Thierry. En effet, le tarif appliqué dépend du type de la chambre ainsi que de la date de réservation. En conséquence, Thierry a décidé de réserver une chambre avec les caractéristiques souhaitées, mais en réduisant la période de résidence. »

1. Extraire des règles métiers à partir de la description du contexte.
2. Proposer un modèle dimensionnel logique de données³ (*assez complet*) du magasin de données concernant **les revenus de l'hôtel**.

Exercice 4 :

Afin de faciliter l'implémentation des données dans le contexte précédent « Réservation des hôtels », définissez son modèle physique de données⁴.

³ Le modèle logique des données est un intermédiaire entre le *modèle entité-association* et le *modèle physique des données*.

⁴ Le modèle physique de données est constitué de tables relationnelles ainsi que d'attributs typés. Les types de données peuvent varier selon les systèmes de gestion de bases de données.

Exercice 5 :

Afin d'observer son *activité de vente* sur les différents lieux de distribution, l'entreprise de fabrication de meubles YY souhaite mettre en place un système d'information décisionnel sous la forme d'un data mart. Ces lieux de distribution sont renseignés par leur enseigne, surface, adresse, département et région. Les ventes se déclinent en période de mois, d'année et de semestre. Elles aussi sont observées par leur type selon le nombre d'articles, et le chiffre d'affaire.

- a) Construire le modèle en étoile de ce data mart indiquant le fait à observer, les axes d'analyse, ainsi que les mesures.
- b) Étendre le modèle construit afin d'avoir un modèle en flocon.

Exercice 6 :

- a) Dans le but d'analyser les ventes d'une entreprise industrielle donnée, créer un modèle en flocon qui permet de mesurer les ventes à partir des quantités vendues et aux bénéfices, en fonction des ventes réalisées par jour, dans un magasin donné, pour un produit donné. Ce modèle doit nous permettre de pouvoir analyser les ventes par jour, par semaine, par mois et par année. Les magasins peuvent être classés selon leur ville et leur pays.
- b) En réduisant la granularité des hiérarchies des dimensions représentant le temps et la localisation des magasins, déduire le modèle en étoile qui permet d'analyser les ventes pour cette entreprise.
- c) Dans le but de mesurer le nombre de commandes qui est donné par jour et par magasin, étendre le modèle en flocon précédent en un modèle de constellation (ajouter une table de fait nommée « Vente-jour »).

NB. N'oubliez pas que ces modélisations ont des fins analytiques et puis décisionnelles.

Correction 1 :

Pour passer d'un modèle conceptuel à un modèle logique, il faut penser à appliquer les transformations suivantes :

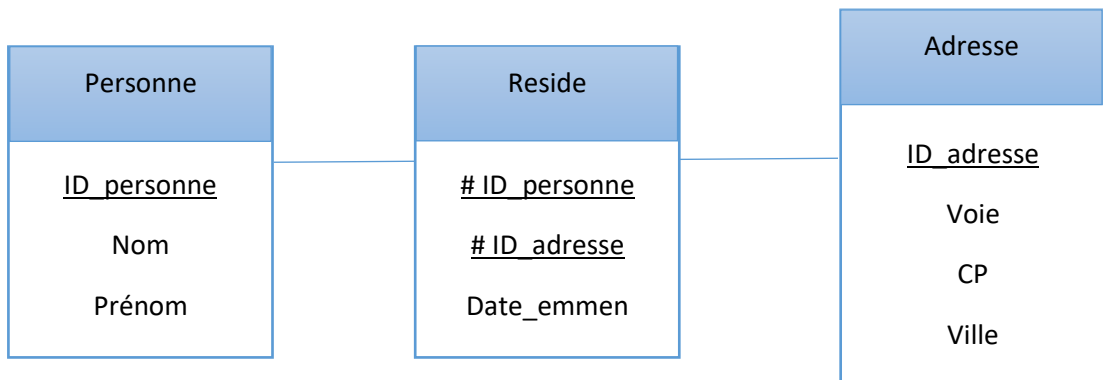
- Entité → Table
- Identifiant → Clé primaire
- Association 1 à 1 → Clé dans la « table à 1 »
- Association 1 à plusieurs → Clé étrangère dans la « table à 1 »
- Association plusieurs à plusieurs → Table supplémentaire avec deux clés étrangères
- Association n-aire → Table supplémentaire avec n clés étrangères

Les solutions pour les modèles de conception proposés sont alors les suivants :

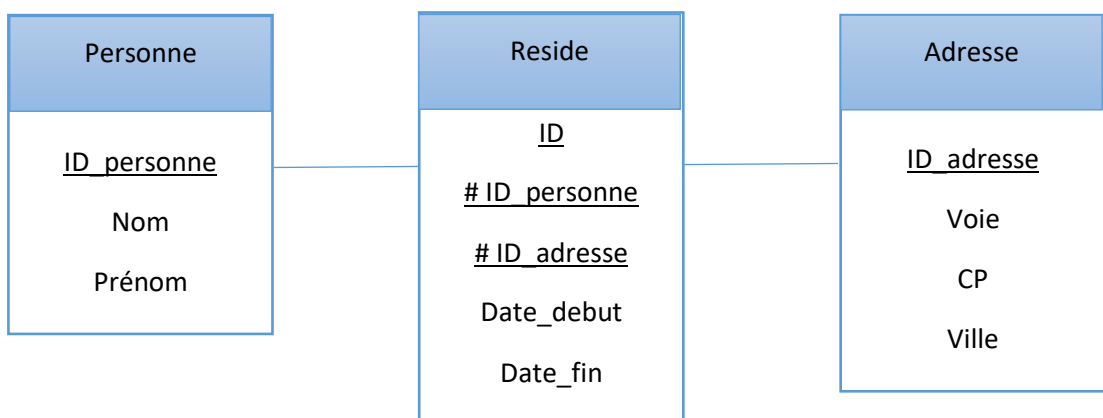
1. Modèle 1 :



2. Modèle 2 :

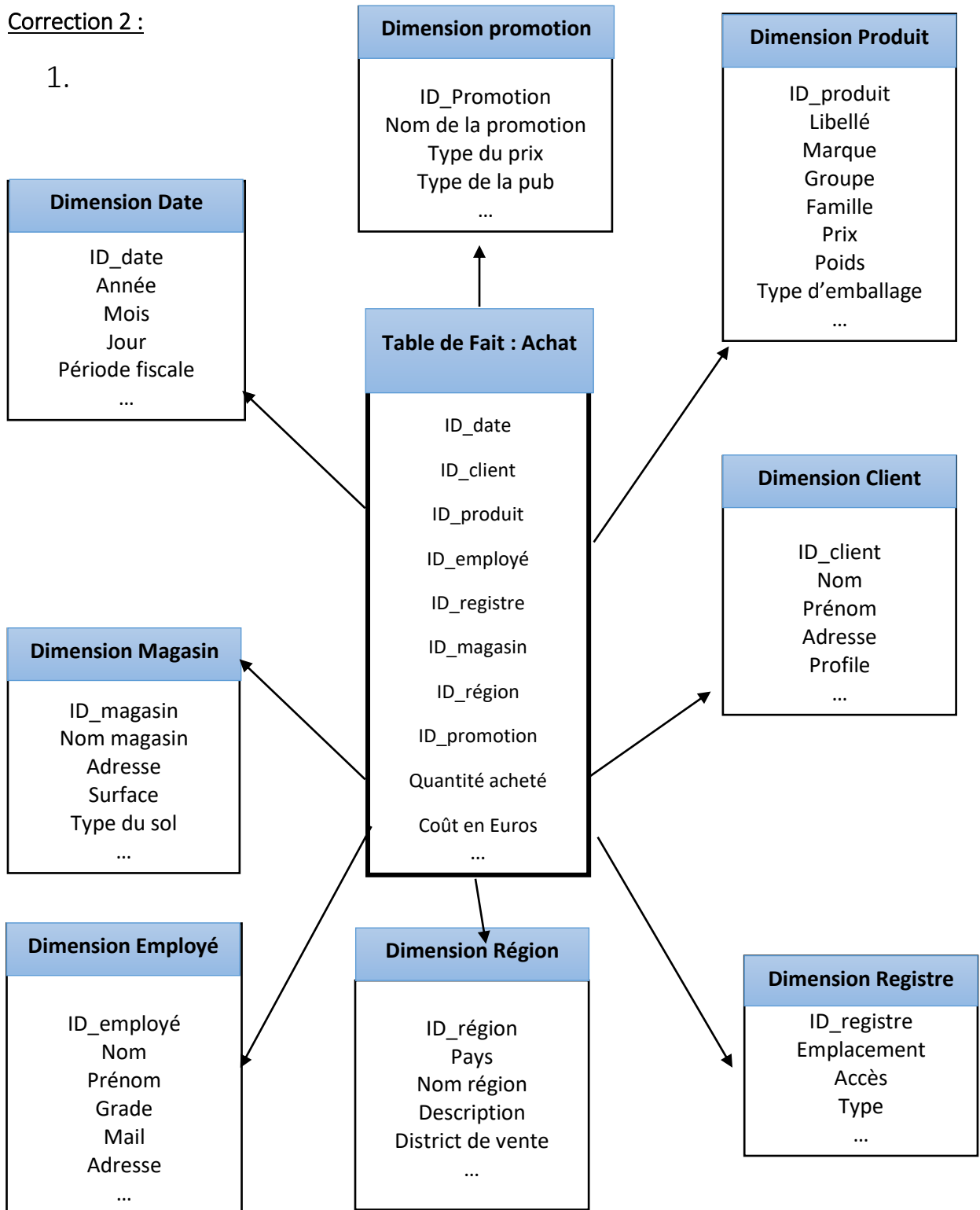


3. Modèle 3 :

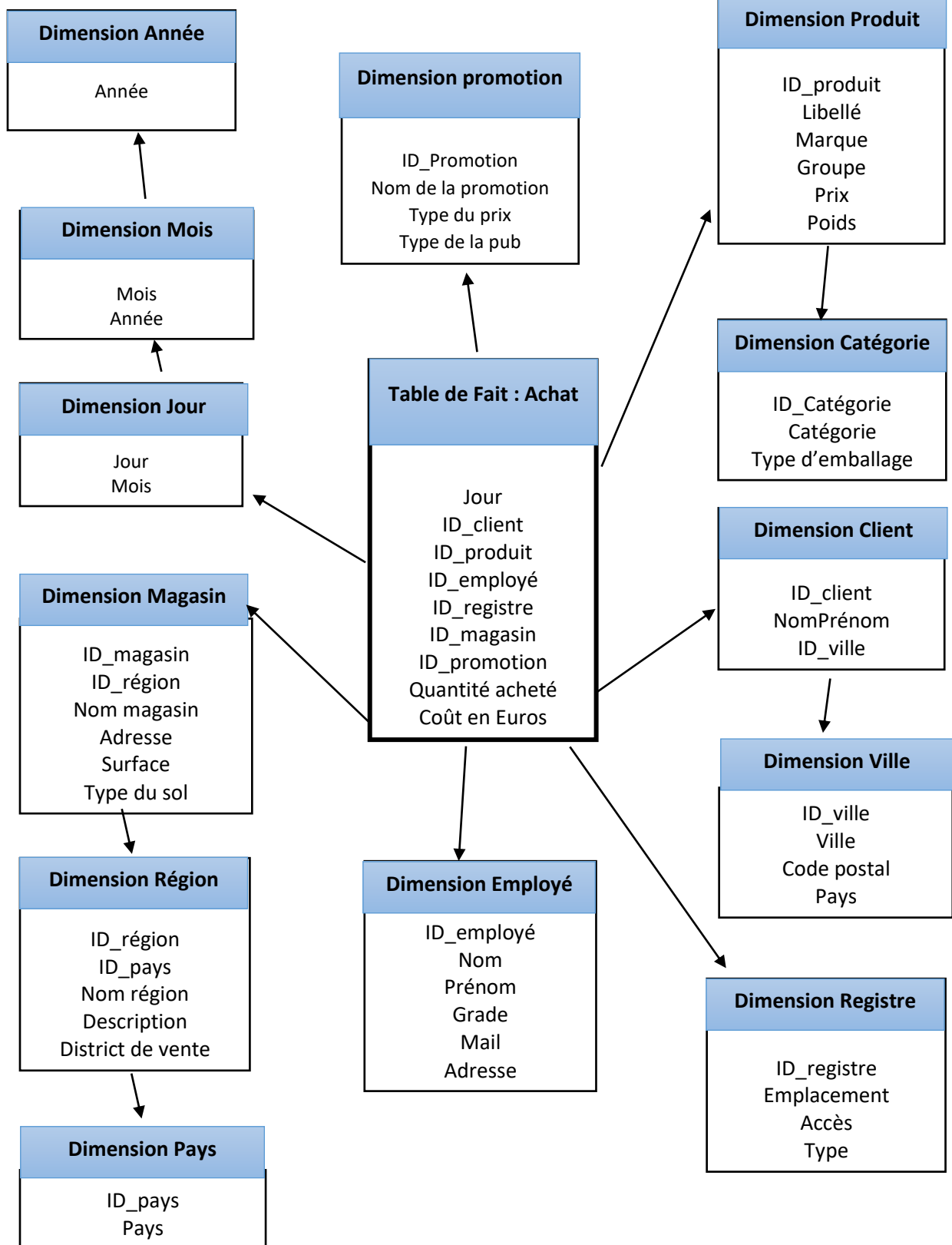


Correction 2 :

1.



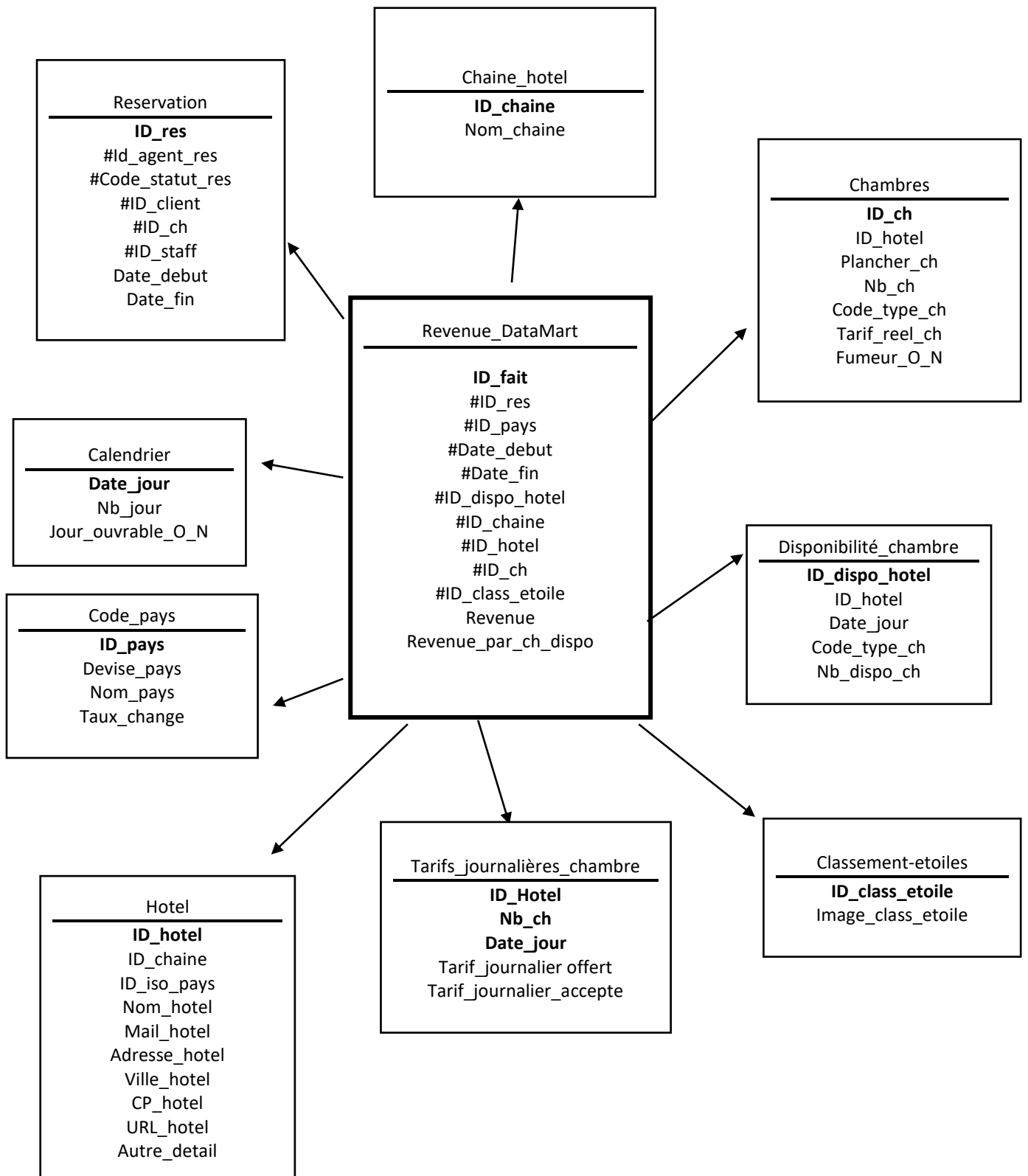
2.



Correction 3 :

1. Règles métier :
 - Les hôtels peuvent (ou non) faire partie d'une chaîne hôtelière
 - Chaque hôtel est dans un pays spécifique
 - Chaque hôtel a un classement par étoiles.
 - Un hôtel peut avoir de nombreuses caractéristiques (comme une salle de sport et un accès Internet).
 - Une caractéristique peut être trouvée dans de nombreux hôtels.
 - Les hôtels ont un certain nombre de chambres d'hôtel de types désignés.
 - Chaque type de chambre a un tarif qui s'applique pour une période de temps spécifiée.
 - Les clients font des réservations
 - Les réservations deviennent des réservations de chambres spécifiques plus tard, généralement lorsque les clients arrivent à l'hôtel.

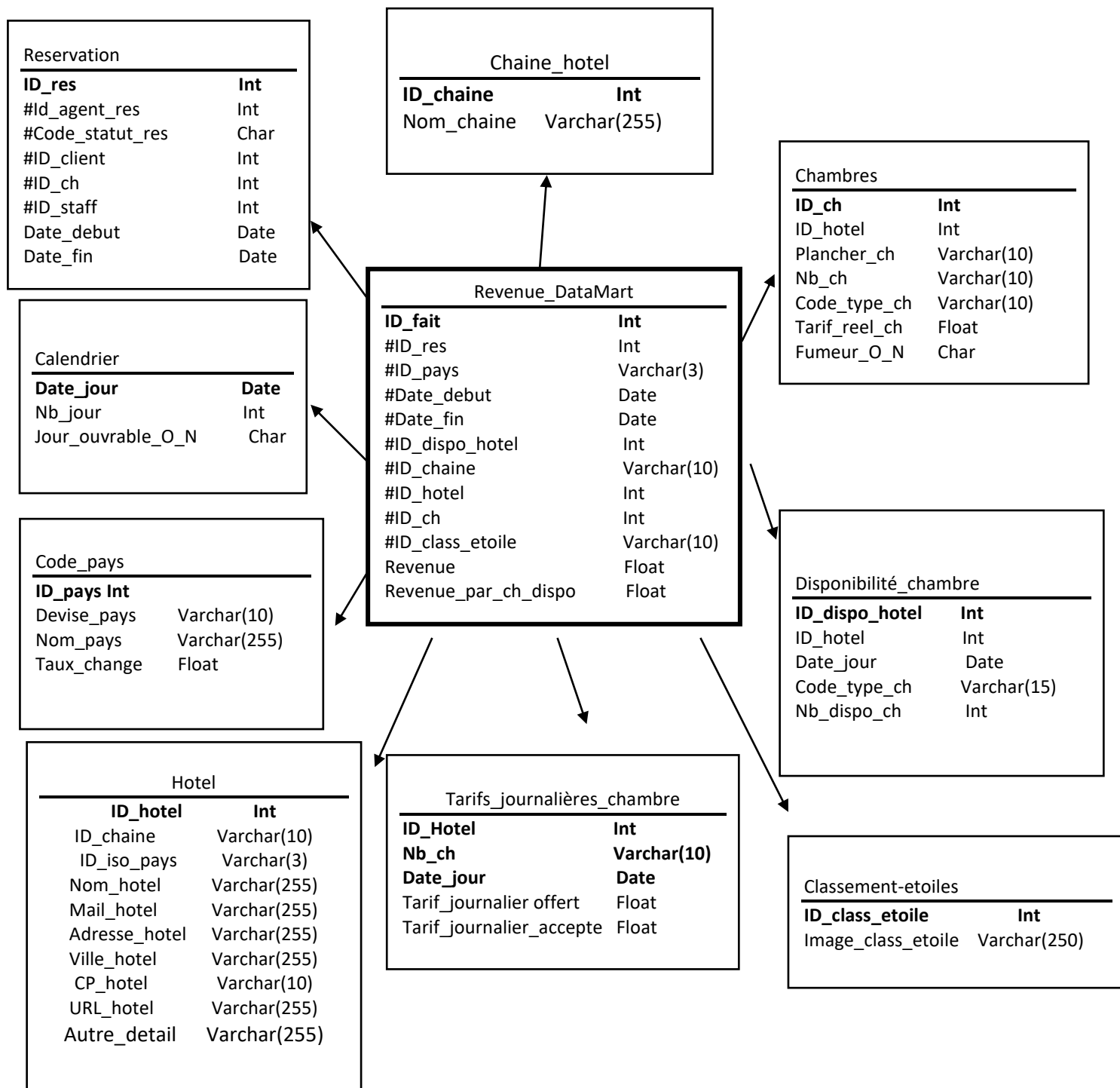
2. La modélisation dimensionnelle logique des données liées au contexte « Réservation des hôtels » peut être proposée comme suit :





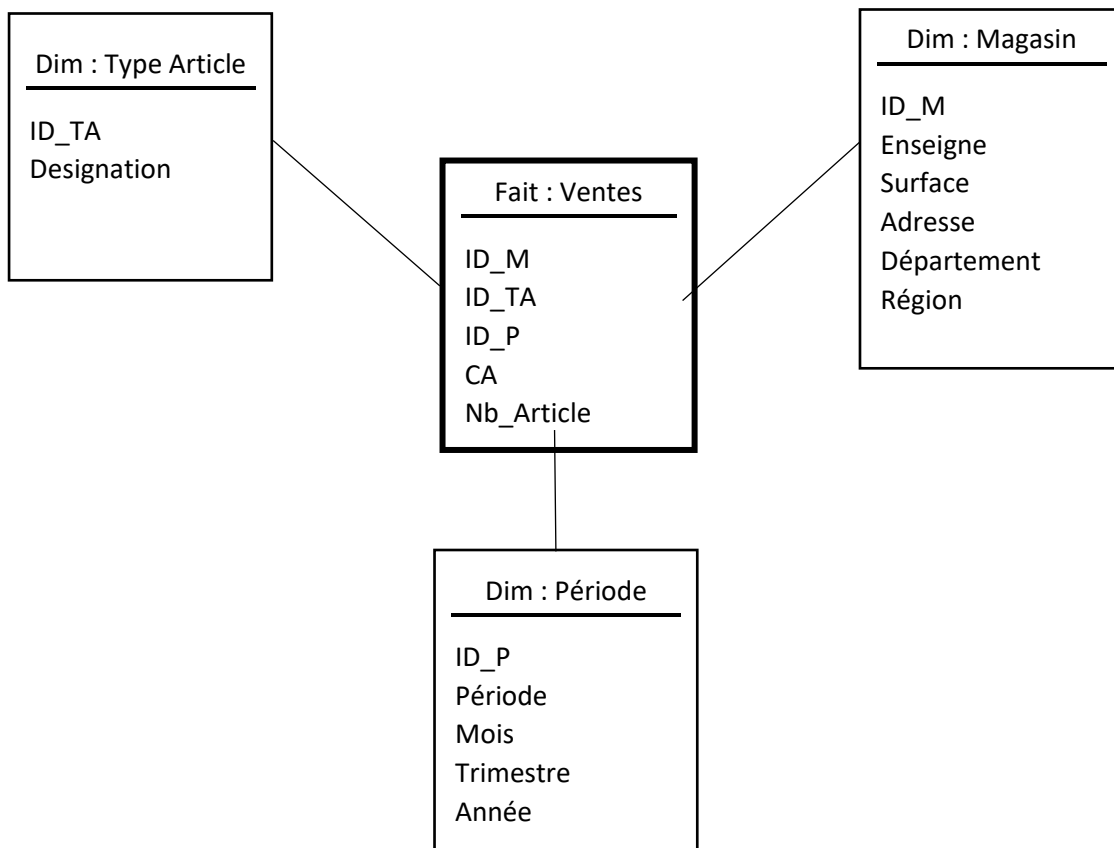
Correction 4 :

Le modèle physique de données qui correspond au contexte « Réservation des hôtels » est le suivant :

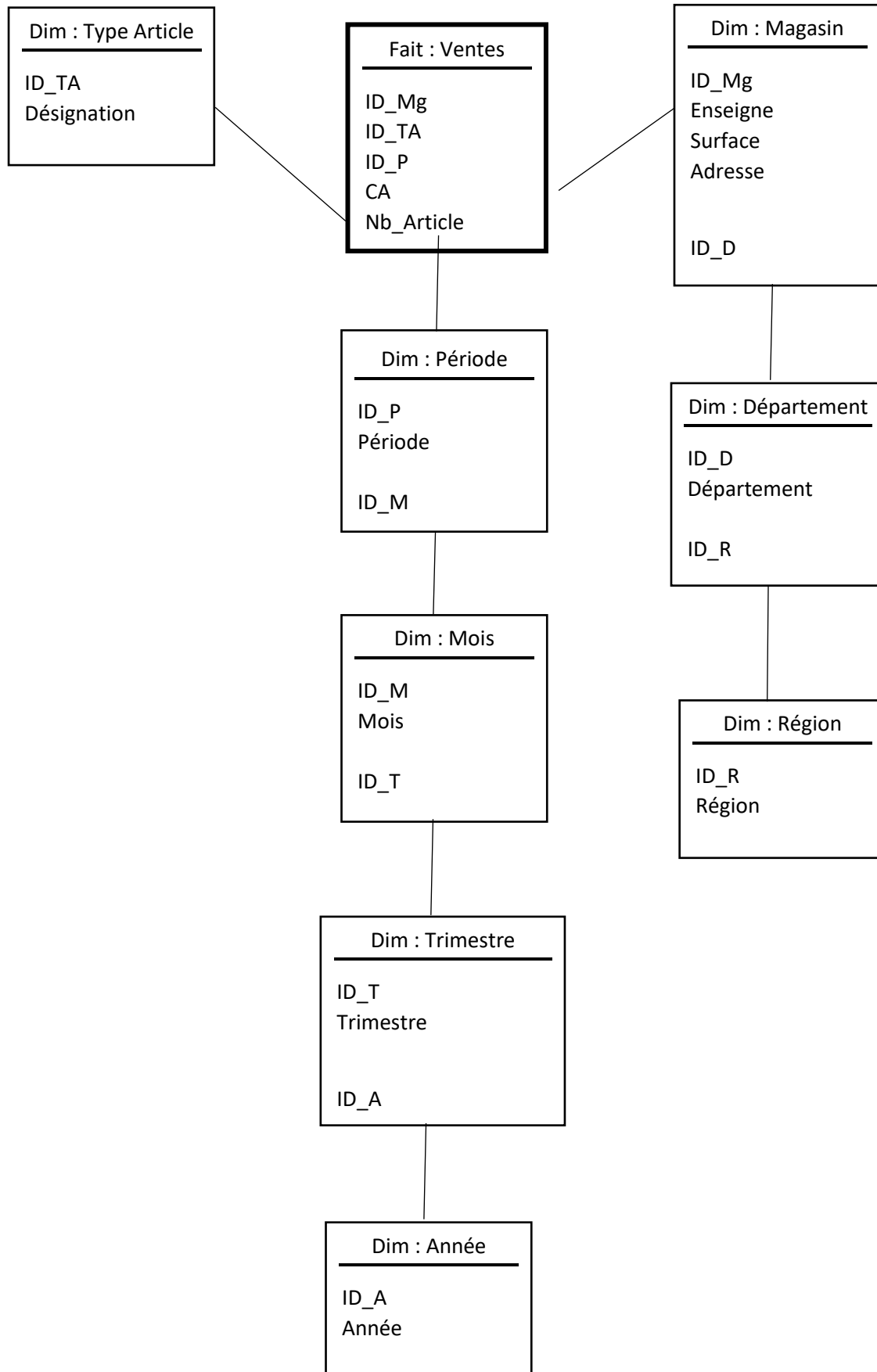


Correction 5:

a. Modèle en étoile :



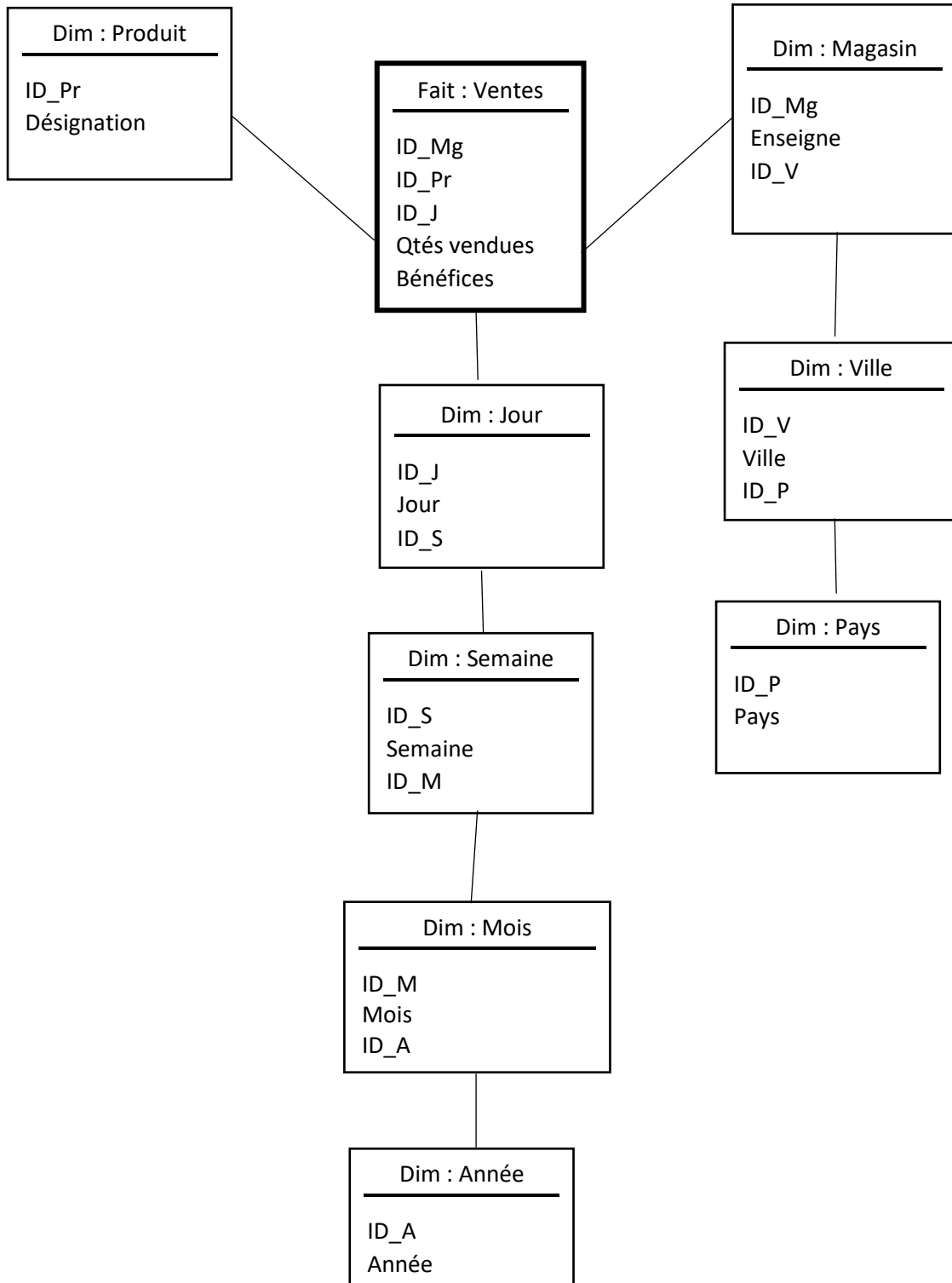
b. Modèle en flocon :



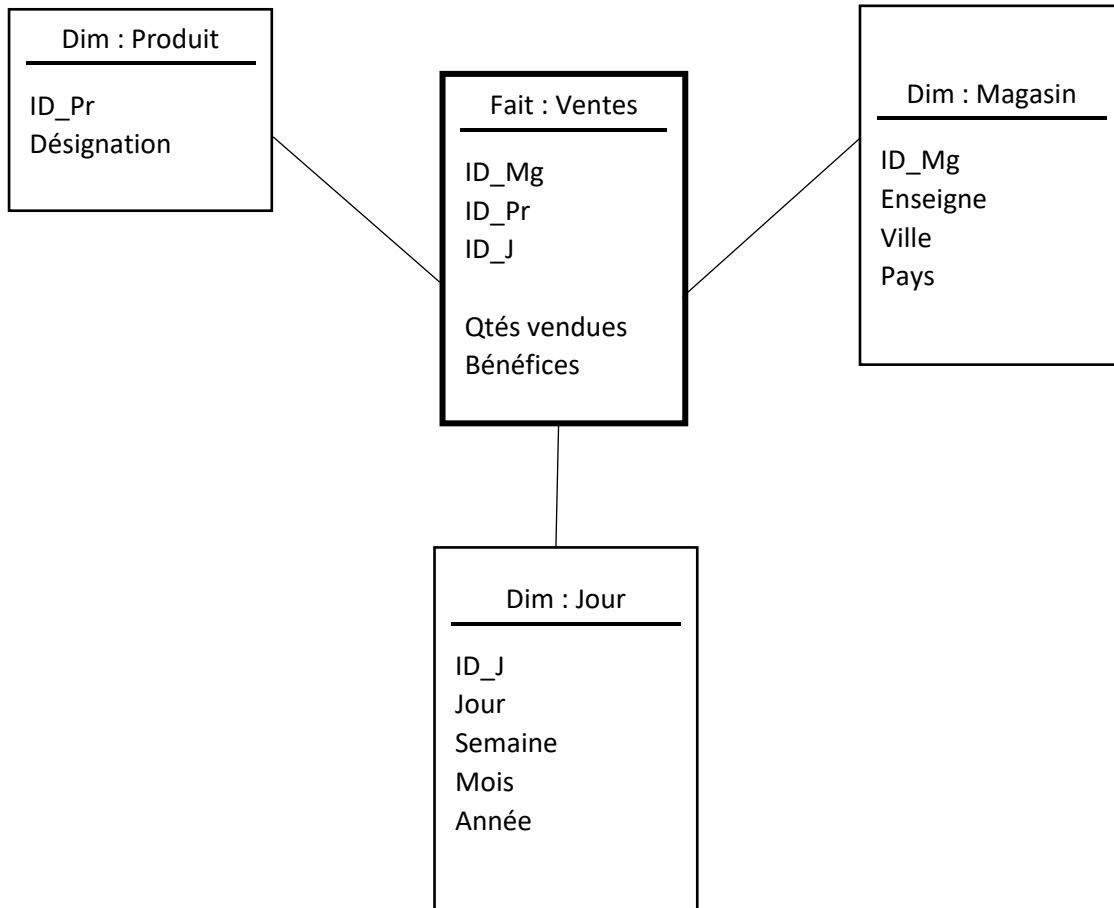
S

Correction 6 :

a. Modèle en flocon :



a. Modèle en étoile :



a. Modèle en constellation :

