

## TD Business Intelligence BI

### LST IGI 2013/2014

#### Exercice 1

Les chiffres d'affaires réalisés par une entreprise. Faire une représentation du cube OLAP.

#### Année n-1

Produit	Autres	Particuliers	professionnels
Informatique	6	10	4
Meubles	6	4	2
Photo	4	7	6

#### Année n-2

Produit	Autres	Particuliers	professionnels
Informatique	8	5	7
Meubles	10	3	8
Photo	6	8	5

#### Année n-3

Produit	Autres	Particuliers	professionnels
Informatique	4	8	2
Meubles	10	9	4
Photo	8	9	1

#### Exercice 2 Conception d'un entrepôt de données

Il s'agit de modéliser l'entrepôt de données (ED) des ventes d'une entreprise commerciale. Cette entreprise vend des produits regroupés par familles de produits. Une vente correspond à un produit et un seul ; la vente est effectuée par l'un des vendeurs du service de vente spécialisé dans le produit. La semaine de vente est le numéro de semaine dans l'année. L'ED doit pouvoir fournir le chiffre d'affaires des ventes d'un produit, par date, client, et vendeur, ainsi que toutes les sommes possibles de chiffre d'affaires.

Les objets de l'ED sont les suivants :

**1) produit**, caractérisé par : **code\_produit, code\_famille**

**2) client**, caractérisé par : **code\_client, nom, csp** (catégorie socio-professionnelle)

**3) vente**, caractérisée par : **date, code\_produit, code\_client, code\_vendeur, montant\_de\_vente**

**4) vendeur**, caractérisé par : **code\_vendeur, nom, code\_service**

**5) date**, caractérisée par : **semaine, mois, année** (la date s'écrit par exemple 20020402)

## Questions

1. **Donner les définitions** des quatre termes suivants : table de faits, table de dimension, indicateur, hiérarchie.

2. **Schéma en étoile** : tracer le schéma en étoile dimensionnel de l'ED, en précisant pour chaque table sa nature dimensionnelle (table de faits ou table de dimension), ses clés, ainsi que la nature des champs.

3. **Cube de données** : Dans l'exemple traité, et représenté par le schéma en étoile, le cube de données est un hyper cube à 4 dimensions : **produit, client, vendeur, date**.

Graphiquement, on peut dessiner en perspective 4 types de cubes à 3 dimensions. Définir les 4 types. à quoi correspond chaque type?

4. Supposons un cube D représentant une **coupe** de l'hypercube à 4 dimensions, selon une valeur de la variable **date**. Combien de tableaux à deux dimensions on peut obtenir à partir de D. En tout combien y a-t-il de types différents de tableaux à deux dimensions ?

### Exercice 3 Entrepôts de données

Une agence de voyage aimerait pouvoir analyser ses données afin de planifier de meilleures campagnes de promotion auprès de ses clients. Plus particulièrement, elle aimerait analyser le nombre et le montant des ventes en fonction:

- De la destination: hôtel, ville, pays, région, catégorie de région (ex: bord de mer, alpine, etc.), catégorie de destination (ex: familial ou non), catégorie hôtel (ex: 1- 4 étoiles) ;
  - De la date d'achat: jour de l'année, jour de la semaine, mois, année, saison touristique (ex: basse ou haute saison);
  - De la date de départ: jour de l'année, jour de la semaine, mois, année, saison touristique (ex: basse ou haute saison);
  - Du forfait: nombre de personnes, nombre de nuits, type de forfait (ex: tout inclus, repas inclus, etc.), type de chambre (ex: standard, suite, penthouse, etc.) ;
  - Du client: groupe d'âge, sexe, adresse, type d'acheteur (ex: nouveau, récurrent, etc.) ;
  - Du canal de vente: catégorie (ex: magasin, internet, etc.) ;
  - De la promotion: catégorie (ex: 2 pour 1, rabais 10%, rabais 25%, etc.), début et fin de validité ;
  - Du mode de paiement: catégorie (ex: crédit, comptant, etc.) ;
- a) Proposez un schéma en étoile permettant de faire ces analyses. Identifiez clairement les clés primaires et étrangères des tables de faits et de dimension;
- b) Identifiez, pour chaque table de dimension, une hiérarchie de niveaux de granularité (e.g., attribut1 ← attribut2 ← ...) ;

# Exercice 2 Conception d'un entrepôt de données

---

Il s'agit de modéliser l'entrepôt de données (ED) des ventes d'une entreprise commerciale. Cette entreprise vend des produits regroupés par familles de produits. Une vente correspond à un produit et un seul ; la vente est effectuée par l'un des vendeurs du service de vente spécialisé dans le produit. La semaine de vente est le numéro de semaine dans l'année. L'ED doit pouvoir fournir le chiffre d'affaires des ventes d'un produit, par date, client, et vendeur, ainsi que toutes les sommations possibles de chiffre d'affaires.

Les objets de l'ED sont les suivants :

- 1) **produit**, caractérisé par : **code\_produit, code\_famille**
- 2) **client**, caractérisé par : **code\_client, nom, csp** (catégorie socio-professionnelle)
- 3) **vente**, caractérisée par : **date, code\_produit, code\_client, code\_vendeur, montant\_de\_vente**
- 4) **vendeur**, caractérisé par : **code\_vendeur, nom, code\_service**
- 5) **date**, caractérisée par : **semaine, mois, année** (la date s'écrit par exemple 20020402)

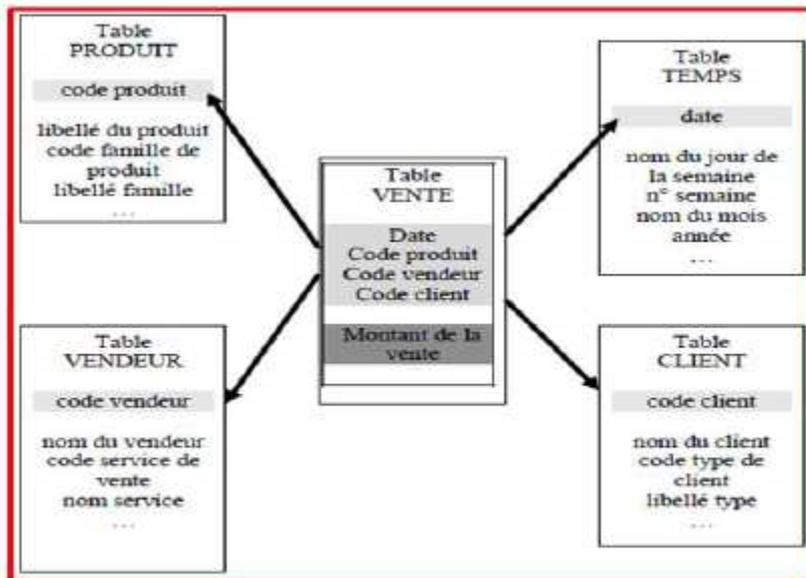
## Questions

1. **Donner les définitions** des quatre termes suivants : table de faits, table de dimension, indicateur, hiérarchie.
2. **Schéma en étoile** : tracer le schéma en étoile dimensionnel de l'ED, en précisant pour chaque table sa nature dimensionnelle (table de faits ou table de dimension), ses clés, ainsi que la nature des champs.
3. **Cube de données** : Dans l'exemple traité, et représenté par le schéma en étoile, le cube de données est un hyper cube à 4 dimensions : **produit, client, vendeur, date**. Graphiquement, on peut dessiner en perspective 4 types de cubes à 3 dimensions. Définir les 4 types. à quoi correspond chaque type?

4. Supposons un cube D représentant une **coupe** de l'hypercube à 4 dimensions, selon une valeur de la variable **date**. Combien de tableaux à deux dimensions on peut obtenir à partir de D. En tout combien y a-t-il de types différents de tableaux à deux dimensions ?

### Correction

#### 1.2



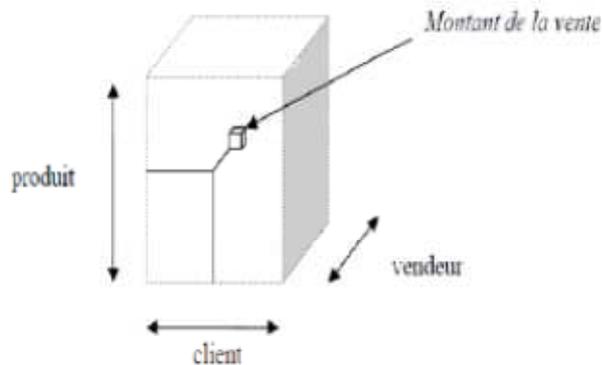
**Table de faits + 4 tables de dimensions**

Ce schéma est caractéristique de la **modélisation dimensionnelle** (du nom des dimensions) la plupart du temps mise en oeuvre dans la conception d'un ED. Un schéma en étoile peut également être représenté sous forme de **cube de données**.

A partir de la base de données relationnelle figurée par notre schéma en étoile, il est possible de développer un logiciel simple (à base de SQL par exemple) qui édite des sommes de « montant de la vente », ou chiffres d'affaires (CA).

3.

Tracé de cube D :



Graphiquement, on peut dessiner en perspective 4 types de cubes à 3 dimensions :

- A. **client, vendeur, date** (pour chaque valeur de produit)
- B. **produit, vendeur, date** (pour chaque valeur de client)
- C. **produit, client, date** (pour chaque valeur de vendeur)
- D. **produit, client, vendeur** (pour chaque valeur de date)

Dans chaque cube, l'élément de base est l'**indicateur** « montant de la vente ».

4. On peut tracer autant de cubes D qu'il y a de valeurs pour la variable **date**.

Un cube D représente une **coupe** de l'hypercube à 4 dimensions, selon une valeur de la variable **date**.

De même, on peut faire des **coupes** du cube D pour toutes les valeurs de **produit**, par exemple. On obtient alors autant de tableaux à 2 dimensions (**client, vendeur**) qu'il y a de valeurs à **produit**.

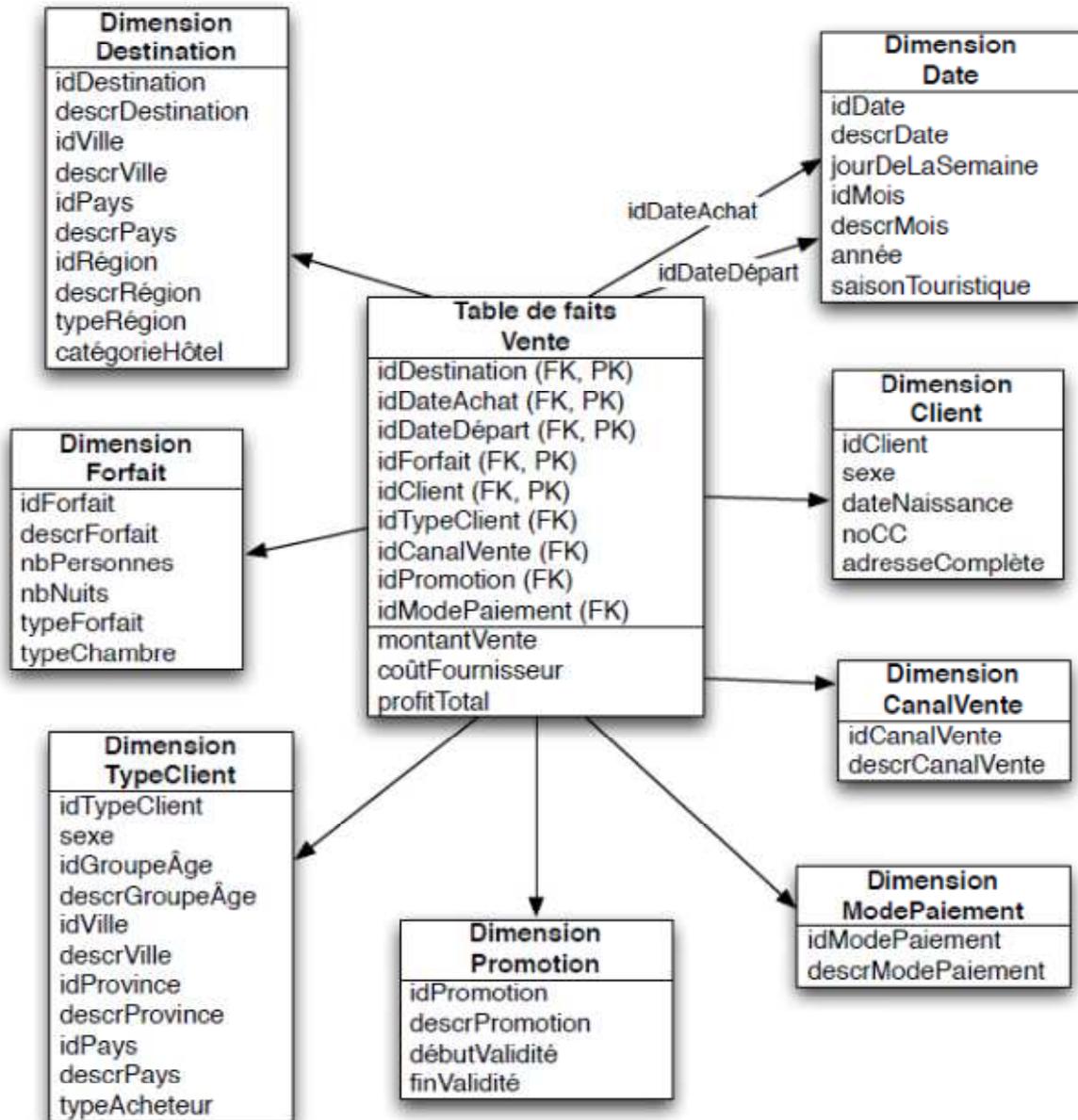
A partir de D, on peut faire 3 types de tableaux à 2 dimensions :

(**client, vendeur**), (**produit, vendeur**), (**client, produit**)

A partir de l'ensemble A, B, C, D, on peut faire en plus les 3 coupes qui gardent **date**

(**client, date**), (**produit, date**), (**vendeur, date**), donc en tout **6 types de tableaux à 2 dimensions**.

### Exercice 3



La plupart des attributs dimensionnels ont un ID ainsi qu'un champ descriptif.

Par exemple, dans la table Date, le mot 'Novembre' n'est pas suffisant pour identifier avec précision ce mois, car on le retrouve dans chacune des années. Il faut donc un attribut idMois (ex: '11/2010') ainsi qu'un attribut descriptif descrMois (ex: 'Novembre'). C'est la même chose pour l'attribut ville: le même nom de ville peut se trouver dans plusieurs pays ou même plusieurs fois dans un même pays;

Nous avons créé une table TypeClient selon la stratégie de mini-dimension.

L'avantage est que la table TypeClient peut être pré-générée (toutes les combinaisons possibles de sexe, ville, groupe d'âge, etc.). De même, les tables Destination, Date, Forfait, Promotion et CanalVentes peuvent également être pré-générée et ne sont (presque) jamais modifiées. Seule la table de dimension Client est modifiée à chaque fois qu'un client s'ajoute au système;

- La clé primaire de la table de faits Vente est une clé composée car il est très rare que l'on accède individuellement les lignes de cette table. En revanche, les clés primaires des tables de dimension sont toujours des clés artificielles simples (ex: NUMBER).

b)

Les niveaux d'une hiérarchie doivent avoir une relation 1 à plusieurs: un parent peut avoir plusieurs enfants (ex: une année a plusieurs mois) mais chaque enfant n'a qu'un seul parent (ex: le mois '11/2010' appartient uniquement à l'année 2010).

Table de dimension	Hiérarchies
Destination	idDestination ← idVille ← idPays ← idRégion ← tous
Date	idDate ← idMois ← année ← tous
Forfait	idForfait ← tous
Client	idClient ← tous
TypeClient	idTypeClient ← idVille ← idProvince ← idPays ← tous
CanalVente	idCanal ← tous
Promotion	idPromotion ← tous
ModePaiement	idModePaiement ← tous