

EXAMEN DE BASES DE DONNEES MULTIDIMENSIONNELLES ET NOMADES

Les deux exercices ne sont pas indépendants, il convient donc d'y répondre dans l'ordre. Chaque exercice compte pour moitié dans la note finale.

Exercice 1. SATISFACTION DE SA VIE

L'Insee réalise une étude sur la satisfaction des Français par rapport à leur vie. L'objectif est de réaliser des tables comme suit :

Satisfaction de sa vie selon le sexe et l'âge en 2021

Tranche d'âge	Femmes	Hommes	Ensemble
16-24 ans	7,1	7,4	7,2
25-34 ans	7,1	7,1	7,1
35-49 ans	7,0	6,9	6,9
50-64 ans	6,5	6,8	6,7
65-74 ans	6,6	6,8	6,7
75 ans ou plus	6,5	6,7	6,6
Ensemble	6,8	6,9	6,8

Note : notation de 0 à 10 ; données provisoires.
Lecture : en 2021, les femmes âgées de 16 à 24 ans évaluent à 7,1 leur satisfaction de la vie qu'elles mènent actuellement.
Champ : France métropolitaine, personnes âgées de 16 ans ou plus.
Source : Insee, dispositif SRCV-Silc.

On souhaite également pouvoir suivre l'évolution de cet indicateur d'année en année, et pouvoir en étudier les variations sur le territoire métropolitain, au niveau national, régional, ou bien départemental. Pour cela, des questionnaires sont envoyés aux sondés, et les résultats sont intégrés à un entrepôt de données. Sur ces questionnaires figurent les questions suivantes :

Questionnaire

Dans quelle tranche d'âge vous situez vous ?

16-24 25-34 35-49 50-64 65-74 75 ou plus

Vous êtes : un homme une femme

Indiquer le code de votre département de résidence : _____

Dans quelle mesure êtes-vous satisfaits de la vie que vous menez actuellement (donnez une note entre 0 et 10) : ____ / 10

Référence du questionnaire : XGFSWERCERE

Dans cet exercice, on va construire un modèle dimensionnel pour le stockage et l'exploitation des réponses à ce questionnaire. Ce modèle sera structuré autour d'une unique table de faits. Les tables de faits et de dimension utiliseront des champs parmi les suivants (en plus des clés primaires et étrangères) :

- Tranche d'âge/sexe/région/département/satisfaction
- référence questionnaire
- jour/mois/année de réception du questionnaire rempli.
-

Question 1. Choisir le nom et le grain de la table de faits.

Nom : faits de satisfaction

Grain : une ligne par questionnaire reçu.

Question 2. Choisir les dimensions

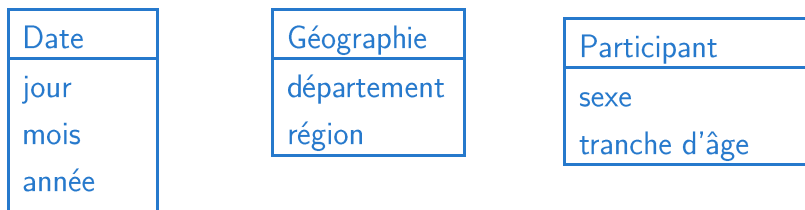
Dimensions :

- Date
- Géographie
- Participant
- Référence questionnaire (dimension dégénérée)

Question 3. Choisir les mesures. Pour chaque mesure, on indiquera (et on justifiera) si elle est additive, semi-additive ou non-additive ?

Une seule mesure : la note de satisfaction. Cette mesure est non-additive, car il s'agit d'un ratio/pourcentage.

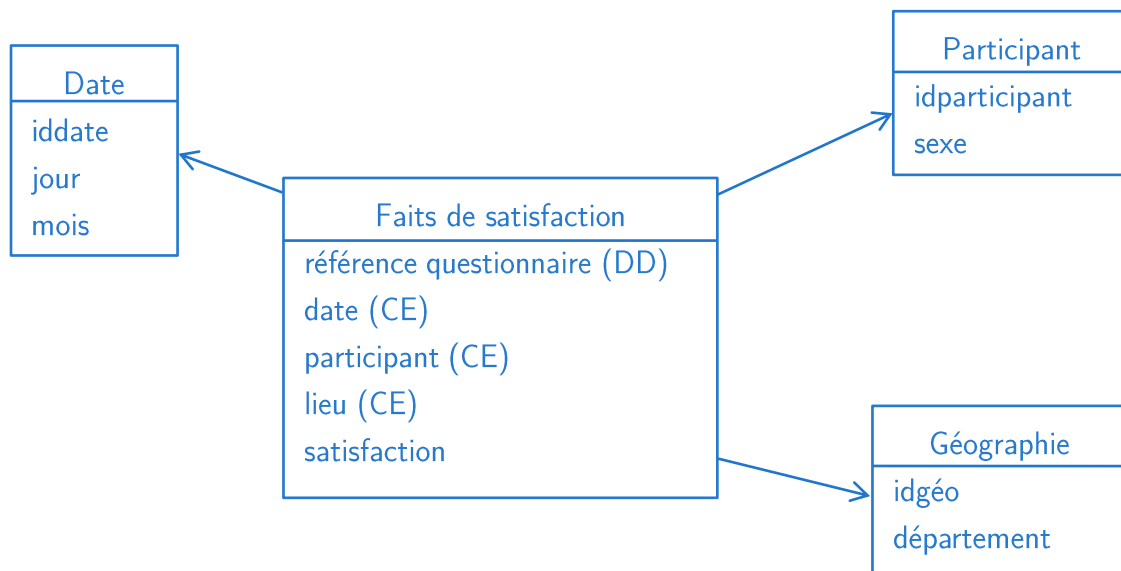
Question 4. Compléter les tables de dimension, en proposant des hiérarchies dimensionnelles.



Hiérarchies :

- jour→mois→année
- département→région

Question 5. Proposer un modèle conceptuel de données complet.



Question 6. (Bonus) Proposer une ou plusieurs requêtes sql permettant de construire la table « Satisfaction de sa vie selon le sexe et l'âge en 2021 » donnée en exemple ci-dessus.

Exercice 2. STRATEGIE DE SONDAGE

Les questionnaires servant à l'étude ci-dessus peuvent être envoyés aux sondés de diverses manières : par courrier, par email, ou bien via des messages sur les réseaux sociaux (Facebook, LinkedIn, Myspace). On souhaite comparer l'efficacité relative de ces différentes méthodes en termes de taux de réponse ainsi qu'en termes de délai de réponse. Pour cela on va introduire dans notre entrepôt une seconde table de faits, enregistrant les envois et les réceptions de questionnaires, ainsi que le mode d'envoi de chaque questionnaire.

Question 1. Choisir le nom et le grain de la table de faits.

Nom : fait de diffusion

Grain : une ligne par questionnaire envoyé ou reçu.

Chaque questionnaire apparaîtra donc une ou deux fois dans la table, suivant qu'il ait été renvoyé au sondeur ou non.

Question 2. Choisir les dimensions, en mettant en avant les dimensions conformes de l'entrepôt.

Dimensions :

- Date (dimension conforme)
- Mode de diffusion (courrier, email, réseau social)
- Référence questionnaire (dimension conforme bien que dégénérée)

Question 3. Choisir les mesures, ou justifier leur absence.

Il n'est pas nécessaire d'avoir des mesures : on peut déduire toutes les valeurs qui nous intéressent à partir de la liste des envois et réceptions de questionnaires, avec les dates et les modes de diffusions.

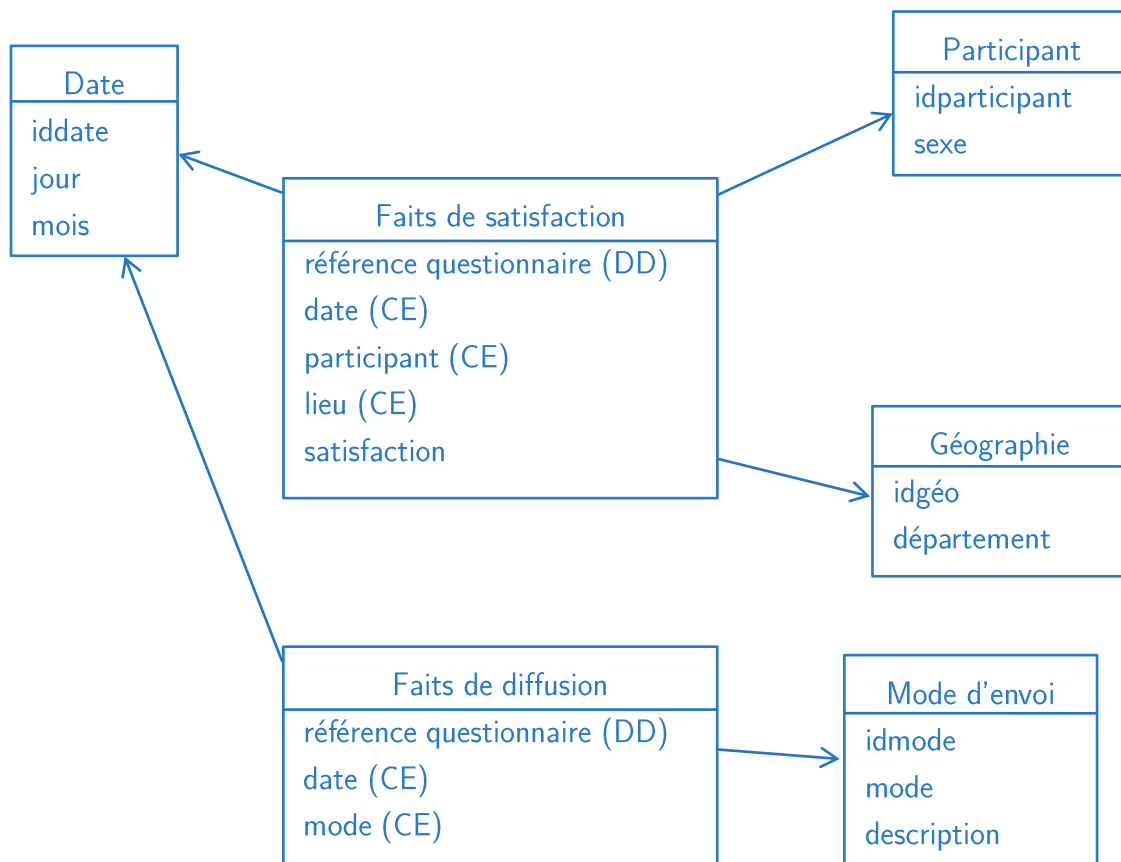
Question 4. Compléter les tables de dimension, et proposer des hiérarchies dimensionnelles.

Date
iddate
jour
mois

Mode d'envoi
idmode
mode
description

Hiérarchie : même hiérarchie que dans l'exercice précédent pour la dimension Date.

Question 5. Proposer un modèle conceptuel de données complet.



Question 6. (Bonus) Proposer une requête permettant de construire la table suivante :

Mode d'envoi	Volume envoyé	Taux de retour	Délai moyen de réponse (en jours)
email	1000000	5%	5
courier	20000	10%	15
Facebook	2000000	45%	6
LinkedIn	10000	60%	2
myspace	5	20%	200