
TP n 14 - Bases de données

Notions abordées

- Comprendre la structure d'une base de données à partir de la description de ses tables
- Formuler des requêtes simples
- Trier et tronquer le résultat d'une requête
- Formuler des requêtes avec agrégation des lignes
- Formuler des requêtes avec jointure de tables.

Exercice 1 Personnes

On reprend la base de données étudiée en classe et contenant les 4 tables :

- `personne`(id_pers, prenom, nom, naissance, pays), représentant une personne par un identifiant unique, son prénom, son nom, son année de naissance et son pays de naissance.
- `agglomeration`(code, nom, nb_habitant, lieu) représentant une agglomération par un code unique, un nom, le nombre de ses habitants, et un pays.
- `habite`(idhabite, id_pers, code_agglo, debut, fin) permettant de représenter qu'une personne d'identifiant `id_pers` a vécu dans la ville `code_agglo` dans les années `debut`, `fin`.
- `capitale`(id_pays, code_agglo) permettant de représenter la capitale d'un pays.

A Requetes élémentaires

Question 1

Donner l'intégralité de la table `personne`.

Question 2

Donner la liste des prénoms des personnes de la table `personne`.

Question 3

Donner la liste des prénoms des personnes de la table `personne`, renommer la colonne `Vorname`.

Question 4

Donner la liste des noms des personnes de la table `personne` et leur année de naissance.

Question 5

Donner la liste des noms des personnes de la table `personne` et leur âge dans une colonne `age` (sachant qu'on est en 2022).

Question 6

Donner la liste des personnes de la table `personne` dont l'âge est de plus de 30 ans.

Question 7

Donner la liste des personnes de la table `personne` nées en France.

Question 8

Donner la liste des personnes de la table `personne` nées avant (au sens large) 1988.

Question 9

Donner la liste des prénoms des personnes de la table `personne` dont l'identifiant est supérieur ou égal à 3 ou qui ont entre 25 et 30 ans.

Question 10

Donner la liste *sans doublons* de tous les pays mentionnés par la base de données.

B Jointures

Question 11

Donner une liste associant les prénoms des personnes de la base avec le code de la capitale de leur pays de naissance.

Question 12

Donner la liste des couples (prénom, code) associant pour chaque personne son prénom et le code d'une agglomération dans laquelle elle a habité.

Question 13

Donner les noms et prénoms des personnes ayant habité à Rome.

Question 14

Donner les noms et prénoms des personnes vivant à Rome en 1999.

Question 15

Donner les noms et prénoms des personnes vivant dans une capitale en 1999.

Question 16

Donner une liste de couples d'identifiants représentant les paires de personnes (différentes) ayant vécu à un moment dans la même ville. (On s'assurera donc que la liste résultat ne contient pas de couple de la forme (i, i) et qu'elle ne contient pas à la fois (i, j) et (j, i)).

C Jointure gauche

Question 17

Donner la liste des noms des agglomérations qui ne sont pas des capitales.

Question 18

Donner, *sans utiliser de fonction d'agrégation*, l'agglomération de plus grande population.

Question 19

Donner, *sans utiliser de fonction d'agrégation*, la différence de nombre d'habitants entre l'agglomération la plus peuplée et l'agglomération la moins peuplée.

Question 20

Donner les noms et prénoms des personnes ayant réussi à fuir le système de surveillance, c'est-à-dire des personnes pour lesquelles il manque les informations d'habitation pendant une période de temps précédent 2022.

Question 21

Donner, *sans utiliser de fonction d'agrégation*, le prénom des personnes ayant vécu dans une ville qui n'est pas une capitale, mais contenant plus d'habitants qu'une des capitales. Avec chaque personne, on donnera aussi la ville en question.

Exercice 2 Jeux de société

Dans cet exercice on s'intéresse à une base de données présentant plusieurs jeux de société. Les données sont scindées en 4 tables, dont vous trouverez une description ci-dessous.

La table jeux. Elle décrit plusieurs jeux de société à travers les colonnes suivantes.

- Une colonne `nom` **TEXT** contenant le nom du jeu.
- Une colonne `nbj_min` **INTEGER** indiquant le nombre minimum de joueurs pour ce jeu.
- Une colonne `nbj_max` **INTEGER** indiquant le nombre maximum de joueurs pour ce jeu.
- Une colonne `nbj_pair` **INTEGER** indiquant si le nombre de joueurs à ce jeu doit nécessairement être pair (1 pour oui ,0 pour non).
- Une colonne `age_min` **INTEGER** indiquant l'âge minimal conseillé pour le jeu.
- Une colonne `duree_th` **INTEGER** (pour durée théorique) indiquant la durée (en minutes) d'une partie mentionné sur la boîte du jeu.

La table parties. Elle liste les parties effectuées en enregistrant pour chaque partie :

- le jeu dans la colonne `jeu` **TEXT** ;
- la date à travers les colonnes `annee` **INTEGER**, `semaine` **INTEGER** et `jour` **INTEGER**, qui indiquent respectivement l'année (nombre à 4 chiffres), la semaine dans l'année (nombre entre 1 et 52), puis le jour de la semaine (nombre entre (1 et 7) ;
- la durée en minutes de la partie dans la colonne `duree_obs` **INTEGER** ;
- l'identifiant de la partie dans la colonne `id` **INTEGER**.

Les tables groupe1 et groupe2. Elles représentent des groupes de joueurs en précisant pour chaque joueur ou joueuse :

- son nom dans la colonne `nom` **TEXT** ;
- son âge dans la colonne `age` **INTEGER** ;
- le temps qu'il ou elle est disponible pour jouer, exprimé en minutes, dans la colonne `dispo` **INTEGER**.

Question 1

Représenter cette base de données par un schéma entité-association.

Question 2

Lister les jeux dont le nombre de joueurs est fixe.

Question 3

Lister les jeux auxquels on peut jouer à 2 joueurs. Combien y en a-t-il ?

Question 4

Lister les jeux auxquels on peut jouer à 5 joueurs. Quelle est leur durée moyenne ?

Question 5

Lister les jeux auxquels peut jouer un enfant de 5 ans.

Question 6

Combien de parties ont été jouées entre 2000 et 2005 ?

Combien de jeux différents ont été joués entre 2000 et 2005 ?

Question 7

D'après la table `parties` quelle est la durée moyenne d'une partie de uno ?

Question 8

Établir les durées moyennes de chaque jeu selon la table `parties`. Présenter ces durées dans une colonne appelée `duree_moy`.

Question 9

Lister les jeux pour lesquels la durée moyenne observée dans la table `parties` affiche un écart relatif à la durée théorique de plus de 10%.

Question 10

Combien de parties ont été réalisées dans les 18 derniers mois (*i.e.* les 78 dernières semaines) ? Disons qu'on est la semaine 21 de l'année 2022. *On pourra d'abord calculer le nombre de semaines passées depuis chaque partie.*

Question 11

À quelle date faut-il remonter pour dire qu'on a fait 10 parties depuis ?

Question 12

À quels jeux peut jouer le groupe de personnes décrit par la table `groupe1` sans tenir compte de leur disponibilité ? Même question pour le groupe décrit par `groupe2`.

Question 13

Même question en tenant compte de leur disponibilité cette fois ? pour le groupe décrit par `groupe2`.

Exercice 3 Transports en commun

Dans cet exercice on s'intéresse au réseau de transport en commun ferroviaire ¹ de la région Île-de-France. Les données sont scindées en 4 bases, dont vous trouverez une description et un extrait ci-dessous.

La table `stations`. Elle représente les points d'arrêts du réseau de transport. Cette table contient 3 colonnes :

¹et les trams

- Une colonne `id` **INTEGER** contenant un identifiant unique de la station.
- Une colonne `nom` **VARCHAR(50)** contenant le nom de la station.
- Une colonne `mode` **VARCHAR(50)** indiquant le type de transport à cette station (Métro, tram, ...).

Plusieurs stations peuvent avoir le même nom, par exemple lorsqu'elles acceptent plusieurs modes de transport.

La table lignes. Elle représente les différentes "lignes" du réseau de transport. Cette table contient 4 colonnes :

- Une colonne `id` **INTEGER** contenant un identifiant unique de la ligne.
- Une colonne `nom` **VARCHAR(50)** contenant le nom de la ligne.
- Une colonne `exploitant` **VARCHAR(50)** contenant le nom de la société exploitant cette ligne.
- Une colonne `mode` **VARCHAR(50)** indiquant le type de transport de cette ligne.

La table arrêts. Cette table fait le lien entre les deux précédentes en donnant les différentes stations où s'arrêtent les différentes lignes. Chaque ligne de cette table représente un arrêt. Cette table contient 3 colonnes :

- Une colonne `station_id` **INTEGER** contenant l'identifiant de la station concernée par l'arrêt.
- Une colonne `ligne_id` **INTEGER** contenant l'identifiant de la ligne concernée par l'arrêt.
- Une colonne `is_terminus` **INTEGER** contenant un entier indiquant si l'arrêt est, ou pas, un terminus.

La table frequentations. Cette table propose quelques données numériques sur la fréquentation de certaines stations. Elle donne le nombre d'entrées depuis la voie publique, par an, pour **certaines** stations. Cette table contient 2 colonnes :

- Une colonne `station_id` **INTEGER** contenant l'identifiant de la station.
- Une colonne `nb_entrees` **INTEGER** donnant le nombre d'entrées par an pour cette station.

id	nom	exploitant	mode
19	13	RATP	Metro
20	7b	RATP	Metro
11	8	RATP	Metro
40	CDGVAL	Transdev	Navette
42	FUNICULAIRE ...	RATP	Navette
9	LIGNE H	SNCF	Train
...

Table lignes

station_id	nb_entrees
41	7574425
862	2995328
...	...

Table frequentations

ligne_id	station_id	is_terminus
19	41	1
19	862	0
9	145	0
...

Table arrêts

id	nom	mode
41	Châtillon-Montrouge	Metro
897	Châtillon-Montrouge	Tramway
862	Gaîté	Metro
417	Mairie de Montrouge	Metro
...

Table stations

A Requêtes simples

Question 1

Donner la liste des noms des lignes de transport.

Question 2

Donner l'ensemble (sans doublons donc) des exploitants.

Question 3

Donner la liste des noms des lignes exploitées par la SNCF.

Question 4

Donner l'ensemble (sans doublons donc) des modes de transports qui sont exploités par la SNCF.

Question 5

Donner l'ensemble des identifiants des stations dont la fréquentation annuelle est entre 1000000 et 2000000.

B Requêtes avec tri

Question 6

Donner l'ensemble des noms des stations de Métro triées par ordre alphabétique.

Question 7

Lister par fréquentations décroissantes les stations dont on sait que la fréquentation annuelle est supérieure à 10000000 entrées. Les stations sont ici données par leur identifiant.

C Requêtes avec agrégation

Question 8

Donner le nombre total d'entrées annuelles comptabilisées sur l'ensemble du réseau. Nommer le résultat `somme`.

D Requêtes avec jointures

Question 9

Donner l'ensemble des noms de stations de métro par ordre décroissant de fréquentations annuelles.

Question 10

Donner sans redondance l'ensemble des exploitants des lignes dont un arrêt est un terminus. Vous devriez obtenir tous les exploitants.

Question 11

Donner l'ensemble des stations de la ligne de métro 13.

Question 12

Donner une requête permettant le calcul d'un tableau contenant uniquement la station de plus haute et la station de plus basse fréquentation parmi celles dont la fréquentation est connue.

Question 13

Donner l'ensemble des noms des stations dont le nombre d'entrées annuelles est supérieur à la moyenne.

E Requetes bonus

Question 14

1. Donner l'ensemble des noms des stations qui sont le terminus d'une ligne de métro (comprendre : il existe une ligne de métro passant à cette station, et dont elle est un terminus).
2. Donner l'ensemble des noms des stations de métro qui ne sont pas seulement un terminus pour les lignes de métro (comprendre : il existe une ligne de métro passant à cette station dont elle n'est pas un terminus).
3. Donner l'ensemble des noms des stations qui sont le terminus d'une ligne de métro mais pas le terminus de toutes les lignes qui s'y arrêtent.

Question 15

1. Donner l'ensemble des noms des stations qui ne sont le terminus d'aucune ligne de métro qu'elles desservent (comprendre : pour toute ligne de métro que dessert cette station, elle n'en est pas un terminus).
2. Donner l'ensemble des noms des stations qui sont le terminus de toutes les lignes de métro qui la desservent.
3. Vérifier qu'aucune station ne vérifie les deux contraintes ci-dessus.
4. Trouver cependant l'unique paire de stations de métro (a) et (b), *ayant le même nom* mais pas le même identifiant² et telle que la station (a) vérifie le point 1 et la station (b) le point 2

Exercice 4 Points et géométrie

Dans cet exercice on s'intéresse à une base de données présentant un ensemble de points colorés du plan. Les données sont toutes contenues dans la table `point`. Chaque point y est décrit par 3 attributs : son abscisse x (un entier), son ordonnée y (un entier), et sa couleur `color` (une chaîne de caractères).

Question 1

Donner la liste des couleurs apparaissant dans la base.

Question 2

Donner pour chaque couleur l'abscisse et l'ordonnée moyennes des points de cette couleur.

Dans la suite de l'exercice on s'intéresse à des quadrilatères particulier formés par des points de la base. On exclut les quadrilatères triviaux, c'est-à-dire constitués de plusieurs fois le même point, de sorte que leur côtés, ainsi que leur surface, sont toujours non nuls. On appelle *rectangle* (resp. *carré*) les rectangles (resp les carrés) dont les côtés sont parallèles aux axes du plan.

²bizarrerie des données qu'il faudrait peut-être signaler à Île de France Mobilités...

A Les rectangles

Question 3

Donner la liste des coordonnées (x_1, y_1) et (x_2, y_2) des paires de points $\{p_1, p_2\}$ telles que $[p_1, p_2]$ forme un segment horizontal de longueur non nulle, avec p_1 à gauche de p_2 .

Question 4

Donner la liste des rectangles décrits par les coordonnées de leurs 4 coins. Combien y en a-t-il si on considère que deux rectangles sont égaux dès lors qu'ils ont les 4 même points

Question 5

Donner la liste des rectangles décrits par les coordonnées de leur coin supérieur gauche, leur largeur (au sens de longueur des côtés horizontaux) et leur hauteur au sens de longueur des côtés verticaux).

Question 6

Donner la liste des rectangles (représentés au choix comme dans l'une des deux questions précédentes) dont les coins sont tous de la même couleur, et préciser cette couleur.

Question 7

Donner la surface moyenne des rectangles.

Question 8

Donner l'ensemble des longueurs des côtés des carrés.

B Les losanges

On appelle *losange* les losanges dont les diagonales sont parallèles aux axes du plan. Dans ce contexte un losange est donc caractérisé par un segment horizontal et un segment vertical qui se coupent en leur milieu (car, on le rappelle, les diagonales d'un losange se coupent en leur milieu, sinon c'est juste un "cerf-volant").

Question 9

Lister les losanges monochromes décrits par les coordonnées de ses sommets en commençant par celui du haut et en tournant ensuite dans le sens des aiguilles d'une montre.

Question 10

Lister les losanges monochromes en donnant pour chacun les coordonnées de son centre, sa hauteur (au sens de la longueur de sa diagonale verticale) et sa largeur (au sens de la longueur de sa diagonale horizontale).

Question 11

Donner la surface moyenne des losanges monochrome.

Question 12

Donner la liste des losanges dont le centre est un des points de la base.