

## TD sur les bases de données réparties

### Exercice 1 :

L'administrateur du portail web d'une entreprise spécialisée en vente par internet (e-commerce) dispose d'une table Utilisateur dont le schéma est le suivant :

**Utilisateur (IDU, Nom, Prénom, email, mot\_de\_passe, Date\_Enregistrement, Date\_Dernier\_Accès, Ville, Pays, Code\_Postal, Téléphone, Age).**

Soient les requêtes R1, R2 et R3 suivantes :

- R1 : Select Nom, Prénom, email From Utilisateur where Date\_Dernier\_Accès < '01-03-2011';
- R2 : Select Ville, Pays, Age From Utilisateur where Date\_Enregistrement < '01-01-2011';
- R3 : Select Nom, Prénom, Ville From Utilisateur where Age>18;

**Objectifs de la fragmentation** : Pour optimiser les requêtes R1 et R2, l'administrateur décide de fragmenter verticalement la table **Utilisateur** en trois fragments : User1, User2 et User3. Ces fragments sont définis pour que chaque requête charge uniquement les colonnes nécessaires à son exécution (R1 charge USER1, R2 Charge USER2).

Donnez les expressions algébriques permettant de représenter chaque fragment

- a. Fragment USER1:
- b. Fragment USER2:
- c. Fragment USER3:

Donnez la réécriture des requêtes R1 et R2 sur la table Utilisateur fragmentée. La réécriture de la requête passe par un processus d'optimisation suivant:

- Dessiner l'arbre global (en considérant la relation avant fragmentation).

- Dessiner l'arbre canonique: remplacer la relation globale par la jointure des fragments pour une fragmentation verticale ou bien l'union dans le cas de la fragmentation horizontale.

## Exercice 2 :

Trois universités sénégalaise (UGB, UCAD, UADB) ont décidé de mutualiser leurs bibliothèques et leur service de prêts, afin de permettre à l'ensemble des étudiants d'emprunter des ouvrages dans toutes les bibliothèques des universités participantes.

Par exemple, un étudiant de UGB pourra emprunter des ouvrages à la bibliothèque de UCAD ou UADB.

La gestion commune des bibliothèques et des emprunts est effectuée par une base de données répartie, dont le schéma global est le suivant :

- EMPLOYE (**IdEmp**, nom, adresse, statut, Bibliothèque)
- ETUDIANT (**IdEtud**, nom, adresse, université, spécialité, nbreEmprunts)
- OUVRAGE (**IdOuv**, titre, #Auteur, éditeur, année, domaine, stock, site)
- AUTEUR (**IdAut**, nom\_auteur)
- PRET (#Ouv, #Etud, date\_emprunt, date\_retour)

L'attribut *Auteur* est une référence à *idAut* de la relation *Auteur*.

L'attribut *domaine* permet de classer les ouvrages en catégories (physique, maths, informatique, médecine, ...).

L'attribut *stock* désigne le nombre d'ouvrages restant disponibles au prêt.

L'attribut *site* indique la bibliothèque qui gère cet ouvrage.

L'attribut *Ouv* est une référence vers *idOuv* de la relation *Ouvrage*.

L'attribut *Etud* est une référence vers *idEtud* de la relation *Etudiant*.

La gestion de cette application s'appuie sur les hypothèses suivantes :

- un employé est affecté à un seul site
- un étudiant est inscrit dans une seule université, mais peut emprunter dans toutes les bibliothèques.
- un ouvrage emprunté dans une bibliothèque est rendu dans la même bibliothèque.

- Le champ nbreEmprunts de la relation ETUDIANT est utilisé pour limiter le nombre d'ouvrages empruntés simultanément par un étudiant sur l'ensemble des bibliothèques. Il est mis à jour lors de chaque emprunt et chaque retour, quelle que soit la bibliothèque d'emprunt.
- Chaque université gère ses propres étudiants
- Chaque bibliothèque gère son personnel et les ouvrages qu'elle détient.

Les relations globales sont fragmentées et réparties sur les différents sites.

1- Donner la définition des différents fragments en utilisant les opérateurs de l'algèbre relationnelle.

2- Allouer chaque fragment au site adéquat, de manière à obtenir une allocation optimale.

## Base de données insee\_deces:

Nous considérons la base de données **insee\_deces** dont le schéma des données est le suivant :

**personne(idp, nom, prenom, datenaiss, lieunaiss, datedeces, lieudeces, age)**

La relation personne stocke les informations relatives aux décès de français entre le 01/01/1970 et

31/12/2019. On dispose pour chaque personne du nom, de ses prénoms, de sa date et lieu de naissance, de sa date et lieu de décès et de son âge au moment du décès. Les lieux de naissance et de décès correspondent à un code commune (code INSEE).

**region(reg, nom, cheflieu, zone)**

La relation region stocke les régions du Sénégal, avec :

- reg est une chaîne de caractères alphanumérique qui identifie une région,
- nom désigne le nom de la région,
- cheflieu référence le code de la commune qui est le chef-lieu de la région.
- zone désigne la zone géographique de rattachement de la région. Ces zones sont définies de la manière suivante :

| Zone | Régions     |
|------|-------------|
| 1    | Saint louis |

2

Dakar

3

Diourbel

N.B. Pour des raisons de simplification, nous ne considérerons pas dans ces exercices les décès des personnes nées en France et mortes à l'étranger, ainsi que les personnes nées à l'étranger et mortes en France.

### **departement(dep, nom, #cheflieu, #reg)**

La relation *departement* stocke les différents départements français, avec :

- *dep* est une chaîne de caractères alphanumérique qui représente le code identifiant le département,
- *nom* désigne le nom du département,
- *cheflieu* référence le code de la commune qui est le chef-lieu du département,
- *reg* référence la région de rattachement du département.

### **commune(com, nom, #dep)**

La relation *commune* stocke les différentes communes du Sénégal telle qu'elles existent en 2022, avec :

- *com* une chaîne de caractères qui correspond au code INSEE de la commune (à ne pas confondre avec le code postal),
- *nom* désigne l'appellation de la commune (en 2022)
- *dep* référence le département de rattachement de la commune.

### **mairie(#codeInsee, cp, nomOrga, nomCom, email, tel, url, adresse, latitude, longitude, dateMaj)**

La relation *mairie* stocke les informations issues de *data.gouv.fr* sur les mairies du Sénégal, avec :

- *codeInsee* correspond au code INSEE de la commune associé à la mairie,
- *cp* désigne le code postal de la commune
- *nomOrga* désigne l'appellation courante de la mairie
- *nomCom* désigne le nom de la commune,
- *email* désigne l'adresse mail de contact de la mairie
- *tel* désigne le numéro de téléphone de contact de la mairie
- *url* désigne l'adresse du site web de la mairie
- *adresse* désigne l'adresse postale de la mairie
- *latitude* et *longitude* désignent les coordonnées de géolocalisation de la mairie
- *dateMaj* désigne la date de dernière mise à jour des informations relative à la mairie.

## **Exercice 3**

*L'objectif de cet exercice est de vous familiariser avec le modèle centralisé de la base de données insee\_deces .*

Soit la requête Q1 : "Donner les personnes (idp, nom, prenom) nées dans la région Saint Louis et décédées dans la région 'Diourbel'".

1. Écrire Q1 en Algèbre Relationnelle.
2. Écrire Q1 en SQL.

#### **Exercice 4 :**

L'objectif de l'exercice est d'analyser les contraintes proposées sous forme d'hypothèses pour en déduire une décomposition de la base de données.

On considère les sites de 'Saint Louis', 'Diourbel' et 'Dakar' auxquels on associera respectivement les numéros 1, 2 et 3.

On souhaite proposer une fragmentation de la base **insee\_deces** sur les sites de 'Saint Louis', 'Diourbel' et 'Dakar' dans le respect des contraintes suivantes :

**C1.** Le site de Saint louis' stocke les régions de la zone 1. Le site de 'Diourbel' stocke les régions de la zone 2 et le site de 'Dakar' stocke les régions de la zone 3.

**C2.** Chaque site gère les départements rattachés aux régions qu'il stocke.

**C3.** Chaque site gère les communes rattachées aux départements qu'il stocke.

**C4.** L'âge des personnes est stocké sur le site de 'Dakar'.

**C5.** Chaque site gère les nom et prénoms, date de naissance et lieu de naissance des personnes nées dans une commune qu'il stocke.

**C6.** Chaque site gère les nom et prénoms, date de décès et lieu de décès des personnes décédées dans une commune qu'il stocke.

**C7.** Les données sur les mairies sont nécessaires sur tous les sites. Le site de Diourbel sera en charge des mises à jour des informations susceptibles d'être actualisées (téléphone, email...)

1. A partir des contraintes, spécifier la manière dont vous envisagez de décomposer la base **insee\_deces** (type de fragmentation...).

2. Définir algébriquement les différents fragments permettant d'effectuer une décomposition de la base **insee\_decès** respectant les contraintes **C1.** à **C7.**
3. Votre fragmentation est-elle *correcte* ? Pourquoi ?