

DS BD 3BIM 2020

Sergio Peignier

Janvier 2020

1 Exercice 1 (~ 6 points)

Dans cet exercice nous allons construire une Base de Données de biologie systémique d'un organisme vivant, à partir de sources d'informations variées. Tout d'abord on a des informations génétiques: Un gène est caractérisé par son identifiant, son nom, sa localisation (chromosome, position, brin), et sa séquence. Un gène peut coder pour un(des) ARN messenger(s). Ces éléments sont caractérisés par leur identifiant, leur nom, et leur séquence. Un ARN messenger peut coder pour une protéine, qui se caractérise également par un nom, un identifiant et une séquence. Une protéine peut éventuellement présenter différents domaines fonctionnels. Un domaine porte un nom qui l'identifie de manière unique et dispose également d'un petit descriptif. Une protéine, peut participer à une réaction biochimique en tant que produit, réactif ou catalyseur. Dans ce cas il est important de connaître le nombre de protéines qui est produit ou consommé au cours de la réaction. D'autres composés peuvent également participer dans une réaction. Ces composés sont caractérisés par leur identifiant et éventuellement par un texte descriptif. Une réaction se caractérise par son nom, et son taux et peut participer dans un certain pathway métabolique. Finalement on souhaite inclure dans le modèle données d'expression transcriptomique: On souhaite stocker l'expression moyenne d'un ARN messenger dans un organe donné (un organe est uniquement caractérisé par son nom).

Questions:

- Construire le modèle Entité/Association
- Établir le modèle relationnel associé (détailler les étapes importantes)
- Imaginons que par mégarde vous avez créé une table "Gene" qui ne porte que le nom du gène et son identifiant. Comment pouvez vous corriger votre erreur? (écrire les commandes qui permettent de résoudre le problème / plusieurs solutions sont possibles).
- Afin d'alimenter la table "Domaine" on vous a fourni le fichier .csv dont les premières lignes sont affichées ci-dessous. Donnez les commandes qui

vous permettent de charger les données dans votre BD (plusieurs solutions sont possibles). Et si le séparateur était le symbole "+" que feriez vous?

```
name\tdescription
A123\tthis crazy domain does not even exist
A42\tdomain 42
A666\tblablabla
B0\tthis domain does not exist either
```

2 Exercice 2 (~ 2 points)

On vous informe qu'une première tentative de création de cette BD a été entreprise auparavant. Dans cette première BD on proposait la relation suivante:

```
Protein(name,sequence,domain,id gene,sequence gene,nom gene)
```

Questions:

- Est-ce que la syntaxe qui définit cette relation est correcte?
- Quels sont les possibles problèmes avec cette relation?
- Proposer une solution aux problèmes.

3 Exercice 3 (~ 2 points)

Questions:

- Quelles sont les différences conceptuelles entre les BD SQL et NoSQL?
- Si vous deviez modéliser un réseau d'interactions génétiques quel type de BD choisiriez vous? Pourquoi? (les arguments et leur précision comptent plus que la réponse)
- Citer trois familles de BD NoSQL.

4 Exercice 4 (~ 10 points)

Soit le modèle relationnel suivant:

- Gang(nom:str, création:date, disparition:date)
- Personne(nom:str, prenom:str, id:int, naissance: date, mort:date, prisonnier:bool)
- Crime(victime_id: int, criminel_id: int, id:int, city:str, date:date, type:str)
- Appartenance(personne_id:int, gang_nom: str)

Écrire les requêtes suivantes en SQL et en algèbre relationnelle (si possible, sinon indiquer la raison)

- Toutes les informations sur les gangs actifs
- Nom et date de création des gangs créés avant 1970
- Personnes qui sont membres des gangs créés en 2020.
- Membres des gangs qui ne sont pas des criminels
- Membres des gangs qui ne sont pas des criminels et qui sont en prison
- Nombre de criminels actifs par gang actif
- Nombre de crimes par ville commis par le gang "Hells Angels"
- Membres du gang "Hells Angels" ayant tué au moins un membre du gang "Pagans".
- Types d'activités criminelles commises par les "Hells Angels" depuis leur création jusqu'en 1990, et n'ayant plus été commises depuis.
- Nombre de crimes commis par les "Hells Angels" depuis leur création classé par ordre chronologique
- Nombre des criminels victimes d'un crime.
- Nombre d'assassinats commis par chaque gang en ordre décroissant.
- Noms et prénoms des personnes victimes d'elles mêmes.