

SAé S2.04 : Exploitation de BDD



Etudiants :

- BERTOUT Baptiste
- LAWSON Killian

Questions (Exercice 1)

Combien y a-t-il de lignes dans chaque fichier ? :

- Il y a 271 117 dans le fichier athlete_event.csv. Commande utilisée `wc -l athlete_events.csv`
- Il y a 231 régions dans le fichier noc_regions.csv. Commande utilisée `wc -l noc_regions.csv`

Afficher uniquement la première ligne du fichier athlète :

Commande `head -n 1 athlete_events.csv`

Quel est le séparateur de champs ? :

Le séparateur de champs est une virgule. Nous l'avons vu grâce à la commande `cat`

Que représente une ligne ?

Dans le fichier athlète, une ligne représente les données personnelles des athlètes ainsi que les données sur leurs participation aux JO.

Dans le fichier noc_régions, une ligne représente le NOC, le nom du pays et une certaine note au besoin

Note: NOC = Un code à trois lettres correspond à chaque comité national

Combien y a-t-il de colonnes ?

Il y a 15 colonnes dans le fichier athlete_events.csv. Commande utilisée `cat athlete_events.csv | wc -l`
il y a 3 colonnes dans le fichier noc_regions.csv. Commande utilisée `cat noc_regions.csv | wc -l`

Quelle colonne distingue les jeux d'été et d'hiver ?

Les jeux d'hiver et d'été sont distingués par la colonne "season". Commande utilisée `cat athlete_events.csv | head -n 1`

Combien de lignes font référence à Jean-Claude Killy ?

Il y a 6 lignes qui font références à cet homme. Commande utilisée `grep 'Jean-Claude Killy' athlete_events.csv | wc -l`

Quel encodage est utilisé pour ce fichier ?

Le fichier est encodé en us-ascii. Commande utilisée `file -i athlete_events.csv`

Comment envisagez-vous l'importation de ces données ?

Avec la commande `\copy` de PostgreSQL

Importation (Exercice 2)

Pour importer les données, nous avons créé une table import avec toutes les colonnes requises pour cette importation. Nous avons également fait attention à ne pas utilisé de type inutile. Par exemple nous avons utilisé un type CHAR(6) au lieu d'un type TEXT pour les médailles car le plus grand mot compte 6 caractères.

Nous avons récupéré les données du fichier athlete_event.csv vers la table import avec la commande `\copy import from chemin/athlete_events.csv with (format CSV, delimiter ',', HEADER)`.

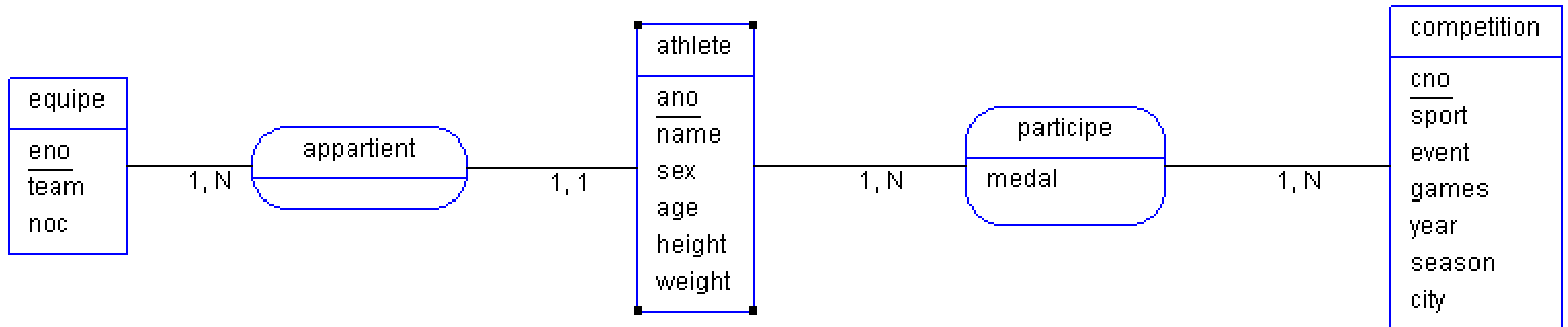
Nous avons donc un total de 271 116 lignes dans la table import. Nous avons supprimé toutes les informations avant 1920 et de toutes les épreuves “Artistiques” car elles étaient incorrectes.

Nous avons également importé les données du fichier noc_regions.csv vers ,une nouvelle table que nous avons créé pour l'occasion, la table regions. Nous avons donc 230 lignes dans cette nouvelle table.

Nous avons finalement fait en sorte que le script soit idempotent, donc qu'il puisse se relancer dans problème. Nous avons donc DROP toutes les tables au début du script.

Ventilation des données (Exercice 4)

Pour ventiler les données nous avons créé 3 tables : une table equipe, une table athlete et une table competition. Nous avons également une table supplémentaire suite à l'association participe entre athlete et competition.



```
equipe (eno, team, noc)
```

```
athlete (ano, name, sex, age, height, weight, #eno)
```

```
competition (cno, sport, event, games, year, season, city)
```

```
participe (#ano, #cno, medal)
```

Personnalisation du rapport (Exercice 6)

Pour cette partie, nous avons choisi de requêter les joueurs et joueuses Français de Handball.

Nous avons donc créé 4 requêtes. Une première requête sur les différentes années auxquelles la France à participer au JO en Handball, une deuxième requête sur le nombre de médailles d'or remportée par la France dans cette discipline. Une troisième requête sur le nombre de femme ayant participé au JO par année. Et enfin une dernière requête sur l'âge moyen des athlètes Français en Handball.

Nous savons donc maintenant que :

- La France a participé à tous les JO de 1992 à 2016.
- La France a gagné 30 médailles d'or dans cette discipline.
- Le nombre de femme ayant participé au JO est de :
 - JO de 2000 → 15
 - JO de 2004 → 15
 - JO de 2008 → 14
 - JO de 2012 → 14
 - JO de 2016 → 16
- L'âge moyen des athlètes est de 27,923 ans