

Rapport BDD

Jeux Olympiques

Nicolas ECKMAN et Aymeri TOURNEUR

Exercice 1

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

```
cat athlete_events.csv | wc -l
```

2. Afficher uniquement la première ligne du fichier athlète

```
cat athlete_events.csv | head -1
```

La première ligne est:

```
"ID", "Name", "Sex", "Age", "Height", "Weight", "Team", "NOC", "Games", "Year", "Season", "City", "Sport", "Event", "Medal"
```

3. Quel est le séparateur de champs ?

- À partir de maintenant, on utilisera une version légèrement modifiée du fichier `noc_regions.csv` en remplaçant les carriage return windows par des retours à la ligne linux et en nommant le fichier `"vnoc_regions.csv"` à l'aide de la commande

```
cat noc_regions.csv | tr "\r\n" "\n" > vnoc_regions.csv
```

4. Que représente une ligne ?

- Pour le fichier `athlete_events.csv`, une ligne représente un athlète, avec ses informations principales comme son nom, son âge, sa taille, les médailles qu'ils ont remportées durant sa carrière, etc...
- Pour `vnoc_regions.csv` une ligne remporte le code du pays (diminutif) ainsi que son nom entier. Il y a aussi une partie, de base vide, qui permet de préciser de quel pays exact il s'agit.

5. Combien y-a t-il de colonnes ?

- Pour `vnoc_regions.csv`:

```
cat vnoc_regions.csv | head -1 | tr ", " "\n" | wc -l
```

- Pour `athlete_events.csv`:

```
cat athlete_events.csv | head -1 | tr "," "\n" | wc -l
```

6. Quelle colonne distingue les jeux d'été et d'hivers ?

```
cat athlete_events.csv | head -1 | tr "," "\n" | nl | grep Games
```

Nous pouvons observer que la colonne 9 est la colonne distinguant les jeux d'été et d'hivers

7. Combien de lignes font référence à Jean-Claude Killy ?

```
cat athlete_events.csv | grep "Jean-Claude Killy" | nl
```

Cette commande nous affiche les 6 lignes mentionnant Jean-Claude Killy

8. Quel encodage est utilisé pour ce fichier ?

```
file -bi athlete_events.csv
```

9. Comment envisagez vous l'import de ces données ?

Nous envisageons un import à l'aide de la commande

```
COPY FROM
```

qui permet de copier d'un fichier vers une table.

Exercice 2

1. Créer une table import permettant l'importation de ces données

```
\i importation.sql
```

2. Remplir cette table avec les données récupérées

```
\copy import FROM 'athlete_events.csv' WITH (FORMAT CSV, HEADER true, NULL  
'NA');
```

3. Supprimez toutes les données strictement inférieures à 1920

```
DELETE FROM import WHERE year < 1920;
DELETE FROM import WHERE sport = 'Art Competitions';
SELECT COUNT(*) FROM import;
```

renvoie 255080

4. Importer tel quel le fichier noc_regions.csv

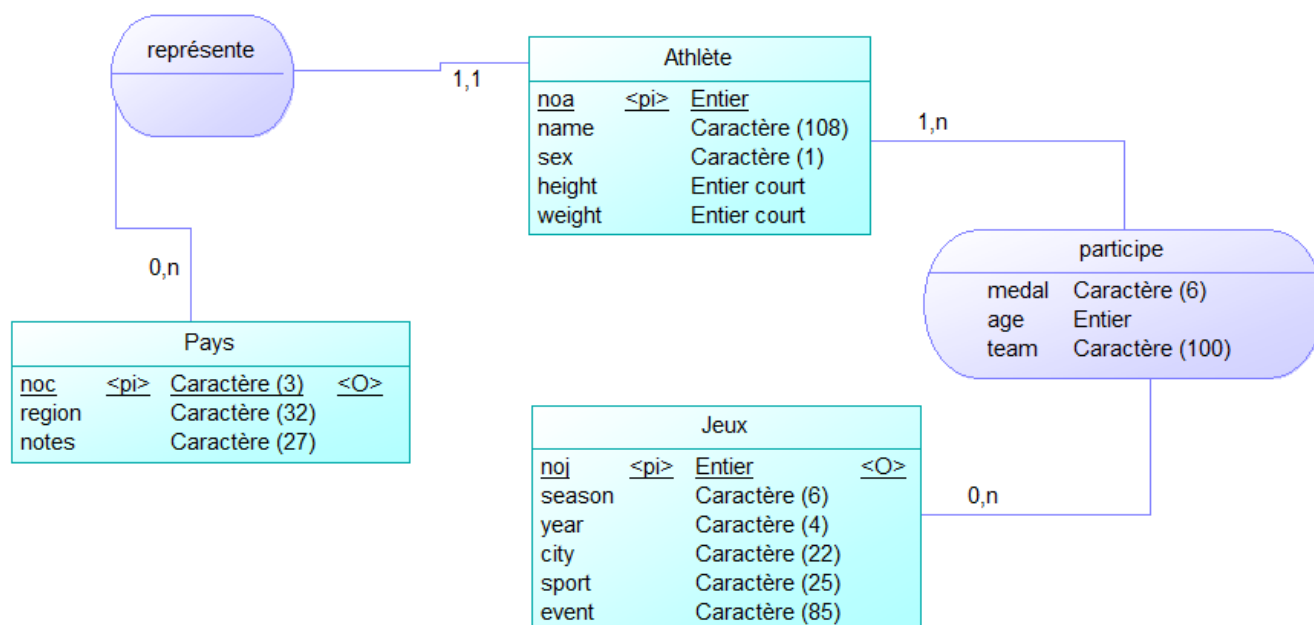
co

Exercice 3

Voir le fichier `requetes.sql`.

Exercice 4

1. Fournir le MCD associé.



2. Fournir le MDL associé.

Pays(noc, region, notes) Jeux(noj, season, year, city, sport, event, #noc) Athlète(noa, name, sex, height, weight) Participe(#noa, #noj, medal, age, team)

- Ajouter dans la table Pays

```
INSERT INTO pays SELECT * FROM noc_regions;
```

- Ajouter dans la table Jeux

```
INSERT INTO jeux(season, city) SELECT DISTINCT season, city FROM athlete_events;
```

2. Une question de taille !

1. Le fichier `athlete_events.csv` fait **41,5Mo**, et le fichier `noc_regions.csv` en fait **3,6Ko**.
2. On peut avoir la taille du fichier avec la commande: `SELECT pg_size_pretty (pg_total_relation_size('import'));`. La table import fait **46Mo**.

Exercice 5

Toutes les requêtes sont dans le fichier `requetes.sql`

Exercice 6



Allemagne

Nous avons choisis le pays Allemagne. Le sport que nous avons pris est la boxe.

Requête n°1 - Le nombre de médaille d'or de l'Allemagne dans le boxe.

Requête n°2 - La ville et le pays où le plus de médaille ont été gagnée par des allemand(e)s dans le sport de boxe. Requête n°3 - Les 3 personnes ayant le plus participé(e) aux épreuves de boxe. Requête n°4 - La taille et l'âge moyen des personnes allemandes ayant gagné(e) une médaille(s) dans le sport de boxe.