
Rapport SAE 2.04

MIREY Kellian, TAS Atilla

1 Compréhension

-Q1.

1-

```
wc -l athlete_events.csv  
=> 271 117 lignes
```

```
cat noc_regions.csv | tr '\r\n' '\n' | wc -l  
=> 230 lignes
```

2-

```
head -n 1 athlete_events.csv  
=> "ID","Name","Sex","Age","Height","Weight","Team",  
"NOC","Games","Year","Season","City","Sport","Event","Medal"
```

3- Le séparateur de champs est la virgule ‘,’.

4- Une ligne représente un athlète avec ses informations.

5-

```
head -n 1 athlete_events.csv | tr ',' ' ' | wc -w  
=> 15 colonnes
```

6- La colonne ‘Season’.

7-

```
cat athlete_events.csv | cut -d ',' -f 2 | grep "Jean-Claude Killy" | wc -l  
=> 6 lignes
```

8-

```
file -i athlete_events.csv  
=> charset = us_ascii
```

9-

```
copy import from 'athlete_events.csv' with  
(format CSV, delimiter ',', HEADER,  
FORCE_QUOTE(Name,Sex,Team,NOC,Games,Season,City,Sport,Event))
```

2 Importation

Après avoir observé la structure du fichier ‘athlete_events.csv’, nous avons décidé de créer la table ‘import’ de la manière suivante en respectant les différents types de données :

```
-- Creation de la table import
Create Table import(
```

```
    ID int,
    Name text,
    Sex char(1),
    Age int,
    Height int ,
    Weight float,
    Team text,
    NOC char(3),
    Games text,
    Year int,
    Season text,
    City text,
    Sport text,
    Event text,
    Medal text
```

```
);
```

Nous avons ensuite importé les données du fichier csv dans la table 'import' à l'aide de la commande suivante :

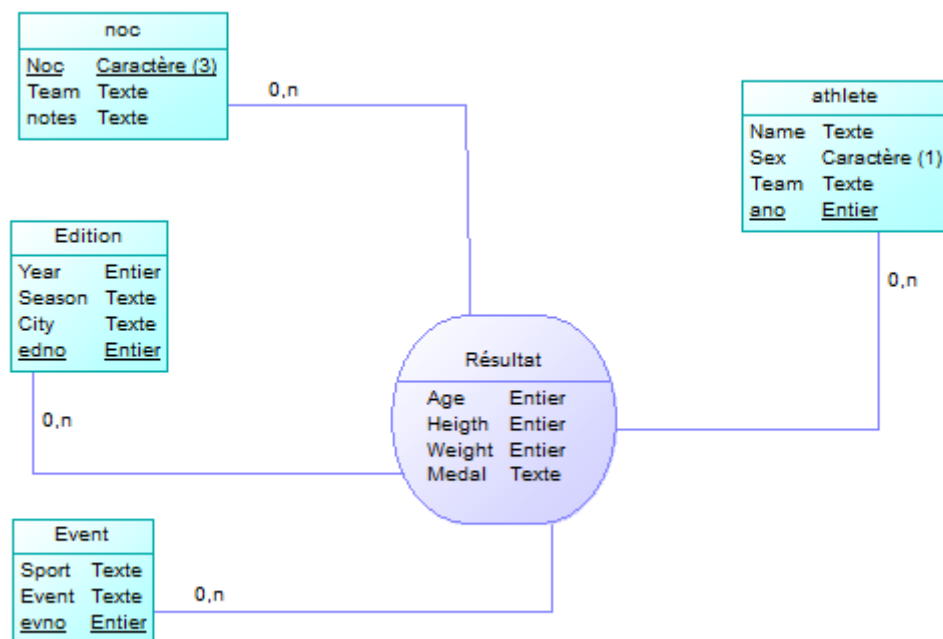
```
\Copy import from 'athlete_events.csv' with csv delimiter ',' header null
↪ as 'NA';
```

Pour retirer les données incorrectes des années antérieures à 1920 et celles des épreuves "Artistiques", nous avons utilisé ces commandes :

```
Delete From import Where year < 1920;
Delete From import Where Sport = 'Art Competitions';
```

3 Ventilation

Pour ventiler les données de la table 'import', nous avons créé 4 nouvelles entités et une association. Les entités représentent les différentes composantes de la table 'import' (ex: athlete) et l'association regroupent ces différentes composantes avec la taille, le poids et la médaille de chaque athlète qui sont des données mesurées lors des épreuves.

**Figure 1:** MCD

3.1 Question

2.1- Le fichier 'data-olympique.zip' fait 5 544 725 octets.

```
wc -c data-olympique.zip
```

2.2- La table 'import' fait 47 259 648 octets.

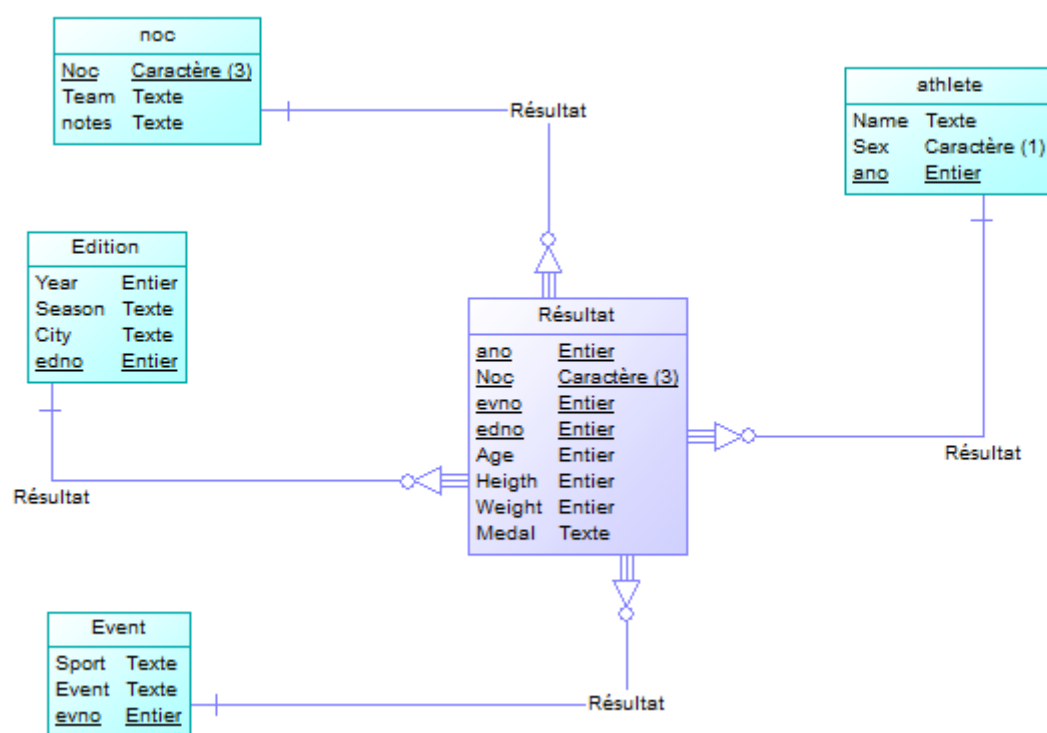
```
Select pg_relation_size('import')
```

2.3- La somme des tables créées fait 9 461 760 octets.

```
Select pg_relation_size('noc')
+ pg_relation_size('athlete')
+ pg_relation_size('edition')
+ pg_relation_size('event')
+ pg_relation_size('resultat');
```

2.4- Nous n'avons pas les droits lors de l'exécution de la commande donc nous n'avons pas le résultat.

```
'Copy resultat to 'resultat.csv' with (format CSV, delimiter ';', HEADER);
```

**Figure 2:** MLD

4 Personnalisation du rapport

Pour cet exercice, nous avons choisi comme pays les États-Unis et comme sport l'athlétisme.

1- Notre première requête affiche le premier et le troisième sur le podium du 200 mètres homme au JO de 1968 au Mexique :

```
Select * From import
Where Sport = 'Athletics'
And Noc = 'USA'
And Year = 1968
And medal in ('Gold','Bronze')
And Season = 'Summer'
And event Like '%200%';
```

Ces athlètes sont Tommie C. Smith et John Wesley Carlos. Ils ont protesté, sur le podium, contre la ségrégation raciale aux États-Unis en lever le poing droit ganté de noir en l'air (symbolisant les mouvements Black Panthers et Black power) et en baissant la tête.

2- La deuxième requête affiche les 10 athlètes ayant le plus de médailles d'or :

```
Select name,sex,Count(*) From import
Where NOC = 'USA'
And Sport = 'Athletics'
And medal = 'Gold'
Group by name,sex
Order by Count(*) desc
Limit 10;
```

On retrouve notamment à la première place Carl Lewis cité plus-tôt dans le sujet.

3- La troisième requête affiche l'évolution du nombre d'athlètes aux USA en athlétisme depuis 1920 à 2016 :

```
Select year,Count(*) From import
Where NOC = 'USA'
And Sport = 'Athletics'
Group by year
Order by year desc;
```

4- La dernière requête affiche la répartition d'athlètes homme et femme durant les JO de 2016 :

```
Select sex, (Count(*)*100)/(Select Count(*) From import
                          Where NOC = 'USA'
                          And Sport = 'Athletics'
                          And Year = 2016) as total
```

```
From import
Where NOC = 'USA'
And Sport = 'Athletics'
And Year = 2016
Group by sex;
```