

---

## **Rapport SAE 2.04: Exploitation de BDD**

MIREY Kellian, TAS Atilla

## Compréhension

### Q1. Analyse du fichier récupéré

Le premier CSV contient des données relatives aux athlètes et aux événements sportifs auxquels ils ont participé, tandis que le deuxième CSV contient des informations sur le comité olympique national (NOC) et la région correspondante.

```
1 1.
2 wc -l athlete_events.csv
3   => 271 117 lignes
4
5 cat noc_regions.csv | tr '\r\n' '\n' | wc -l
6   => 230 lignes
```

```
1 2.
2 head -n 1 athlete_events.csv
3   => "ID","Name","Sex","Age","Height","Weight","Team","NOC","Games","
      Year","Season","City","Sport","Event","Medal"
```

```
1 3. Le séparateur de champs est ``,``.
```

```
1 4. Une ligne représente un athlète avec ses informations.
```

```
1 5.
2 head -n 1 athlete_events.csv | tr ',' ' ' | wc -w
3   => 15 colonnes
```

```
1 6. La colonne `Season` distingue les jeux d'été et d'hivers.
```

```
1 7.
2 cat athlete_events.csv | cut -d ',' -f 2 | grep "Jean-Claude Killy" |
   wc -l
3   => 6 lignes
```

```
1 8.
2 file -i athlete_events.csv
3   => charset = us_ascii
```

Avant de commencer l'importation, il est précisé que les colonnes `Name`, `Sex`, `Team`, `NOC`, `Games`, `Season`, `City`, `Sport`, et `Event` doivent être forcées à être entre guillemets lors de l'importation (`FORCE_Q UOTE (Name, Sex, Team, NOC, Games, Season, City, Sport, Event)`).

```
1 9.
2 copy import from 'athlete_events.csv' with
3 (format CSV, delimiter ',', HEADER,
```

```
4 FORCE_QUOTE(Name,Sex,Team,NOC,Games,Season,City,Sport,Event))
```

## Importation

On commence par supprimer la table import si elle existe déjà (`Drop Table import;`) et à créer une nouvelle table import avec les noms de colonnes appropriés (`Create Table import(...)`).

Ensuite, les données du fichier `athlete_events.csv` sont importées dans la table import en utilisant la commande `Copy import from 'athlete_events.csv' with csv delimiter ',' header null as 'NA';`.

La commande `header` indique que la première ligne du fichier CSV contient les noms des colonnes.

La commande `null as 'NA'` permet de remplacer les valeurs nulles dans le fichier CSV par la chaîne de caractères « NA ».

```
1 1. -- Supression de la table import existante
2 Drop Table import;
3
4 -- Creation de la table import
5 Create Table import(
6     ID int,
7     Name text,
8     Sex char(1),
9     Age int,
10    Height int,
11    Weight float,
12    Team text,
13    NOC char(3),
14    Games text,
15    Year int,
16    Season text,
17    City text,
18    Sport text,
19    Event text,
20    Medal text
21 );
```

```
1 2. -- Importation des données du fichier athlete_events.csv dans la
   table import
2 echo 'Importation des données du fichier athlete_events.csv dans la
   table import'
3
4 Copy import from 'athlete_events.csv' with csv delimiter ',' header
   null as 'NA';
5
```

```
6 -- Affichage du nombre de ligne initial dans la table import
7 Select Count(*) From import;
```

Une fois les données importées, on supprime les lignes où l'année est antérieure à 1920 et où le sport est "Art Competitions".

```
1 3. -- Suppression des données antérieur à 1920
2 Delete From import Where year < 1920;
3 Delete From import Where Sport = 'Art Competitions';
4
5 -- Affichage du nombre de ligne dans la table import
6 Select Count(*) From import;
```

On supprime la table `noc` si elle existe déjà (`Drop table noc Cascade;`) et crée une nouvelle table `noc` avec les noms de colonnes appropriés (`Create Table noc(...)`).

```
1 4. -- Suppression de la table noc existante
2 Drop table noc Cascade;
3
4 -- Creation de la table noc
5 Create Table noc(
6     NOC char(3),
7     region text,
8     notes text Default NULL,
9
10    Constraint pk_noc Primary Key (NOC)
11 );
```

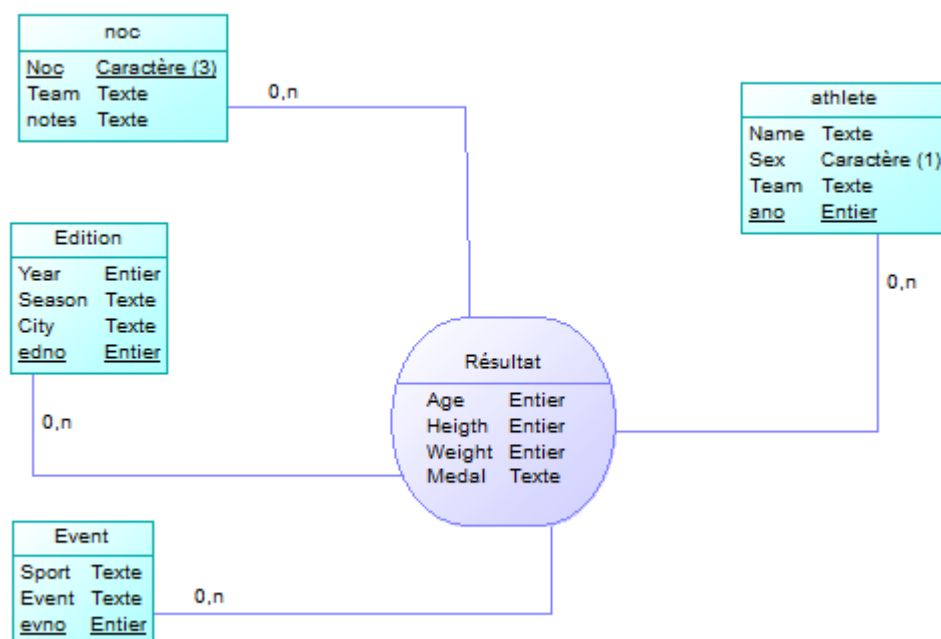
Les données du fichier `noc_regions.csv` sont ensuite importées dans la table `noc` en utilisant la commande `Copy`.

```
1 -- Importation des données du fichier noc_regions.csv dans la table noc
2 echo 'Importation des données du fichier noc_regions.csv dans la table
   noc'
3
4 Copy noc from 'noc_regions.csv' with csv delimiter ',' header;
```

## Ventilation

### Q1. Normalisation des données

On procède à la ventilation des données dans les tables correspondant au MCD. Les tables `athlete`, `edition`, `event`, et `resultat` sont créées en utilisant les commandes `Create Table ...`. Les données sont ensuite insérées dans ces tables à l'aide des commandes `Insert Into ... Select Distinct ... From import` en utilisant les critères de jointure définis.



```

1  --Supression des tables à créer, correspondant au MLD
2  Drop Table athlete Cascade;
3  Drop Table edition Cascade;
4  Drop Table event Cascade;
5  Drop Table resultat Cascade;
  
```

```

1  --Création des tables correspondant au MLD
2  Create Table athlete(
3      ano serial,
4      Name text,
5      Sex char(1),
6
7      Constraint pk_athlete Primary Key (ano)
8  );
9
10 Create Table edition(
11     edno serial,
12     year int,
13     season text,
14     city text,
15
16     Constraint pk_edition Primary Key (edno)
17 );
18
19 Create Table event(
20     evno serial,
21     sport text,
  
```

```
22     event text,
23
24     Constraint pk_event Primary Key (evno)
25 );
26
27 Create Table resultat(
28     ano int,
29     noc char(3),
30     evno int,
31     edno int,
32     age int,
33     heighth int,
34     weight int,
35     medal text,
36
37     Constraint pk_resultat Primary Key (ano,noc,evno,edno,age,height,
38         weight),
39
40     Constraint fk_ano Foreign Key (ano)
41     References athlete(ano)
42     On Update Cascade,
43
44     Constraint fk_noc Foreign Key (noc)
45     References noc(NOC)
46     On Update Cascade,
47
48     Constraint fk_evno Foreign Key (evno)
49     References event(evno)
50     On Update Cascade,
51
52     Constraint fk_edno Foreign Key (edno)
53     References edition(edno)
54     On Update Cascade
55 );
```

```
1  --Insertion des données dans les tables créées précédemment
2  Insert Into athlete (Name,Sex)
3      Select Distinct Name,Sex
4      From import
5      Where Name is not NULL
6      And Sex is not NULL;
7
8  Insert Into edition (year,season,city)
9      Select Distinct Year,Season,City
10     From import
11     Where Year is not NULL
12     And Season is not NULL
13     And City is not NULL;
14
15 Insert Into event (sport,event)
16     Select Distinct Sport,Event
```

```
17      From import
18      Where Sport is not NULL
19      And Event is not NULL;
20
21  Insert Into resultat (ano,noc,evno,edno,age,height,weight,medal)
22      Select ano,noc.noc,evno,edno,age,height,weight,medal
23      From athlete as a, edition as ed, event as ev, noc, import as i
24      Where a.name = i.Name
25      And a.sex = i.Sex
26      And ed.year = i.Year
27      And ed.season = i.Season
28      And ed.city = i.City
29      And ev.sport = i.Sport
30      And ev.event = i.Event
31      And noc.noc = i.NOC
32      And age is not null
33      And height is not null
34      And weight is not null;
```

## Q2. Une question de taille !