

**SAé2.04 semestre 2**

Sujet Olympique

Travail réalisé en groupe de 2 par :

* Antoine DOMISSE
* Antoine POURTHIE

Groupe : B

**Compréhension des exercices :**

**Réponses Exercice n°1 :**

**Convertir un fichier en utf 8 :**

**iconv -f iso-8859-1 -t utf-8 fichier1 > fichier2**

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

Commande :

—> wc -l “nom du fichier”

Pour athlete\_events.csv : 271117 lignes.

Pour noc\_regions.csv : 231 lignes.

1. **Afficher uniquement la première ligne du fichier athlète**

**Commande :**

—> head -n 1 athlete\_events.csv

La résultat obtenu est :

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

1. **Quel est le séparateur de champs ?**

**Commande :**

—> head -n 1 athlete\_events.csv | tr -d [a-z] | tr -d [A-Z] | tr -d [0-9] | tr -d \"

Cette commande permet d’enlever tous les caractères pouvant être présents dans les valeurs des cases. Les guillemets ont aussi été enlevés car chaque valeur est forcément considéré entre guillemets vu que c’est un .csv.

Le séparateur de champ est donc une virgule(‘,’).

1. **Que représente une ligne ?**

**Commande :**

—> sed -n 5p athlete\_events.csv

Une ligne représente les informations d’une personne précise, avec son âge, son nom, son domaine sportif etc.

1. **Combien y-a t-il de colonnes ?**

**Commande :**

—> sed -n 1p athlete\_events.csv | tr , " " | wc -w

Cette commande montre qu’il y a 15 colonnes différentes, on a remplacé la virgule(= le séparateur de champ) par un espace pour que la commande wc -w puisse compter le nombre de mots.

1. **Quelle colonne distingue les jeux d’été et d’hivers ?**

**Commande :**

—> sed -n 6p athlete\_events.csv | cut -d , -f 11

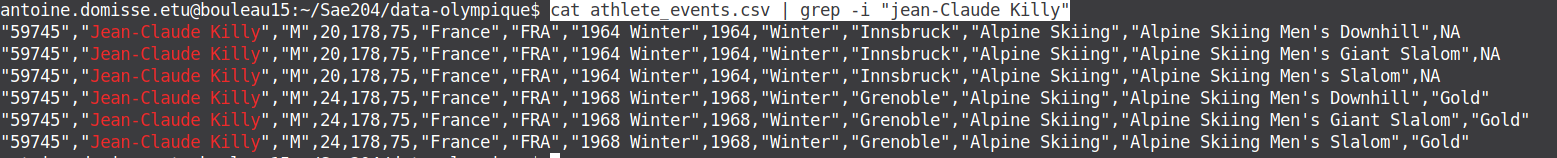
On a décidé d’essayer la commande pour la 5ème personne en prenant la colonne 11 puisqu’on sait qu’elle se nomme “Season” et on obtient comme résultat “Winter” donc cela confirme que c’est bien la colonne 11 qui distingue les jeux d’été des jeux d’hiver.

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

**Commande :**

—> cat athlete\_events.csv | grep -i "jean-Claude Killy"

Cette commande nous renvoie ce résultat :



Donc 6 lignes en font références.

1. **Quel encodage est utilisé pour ce fichier ?**

**Commande :**

—> file athlete\_events.csv

Cette commande nous dit que le fichier est un “CSV FILE”.

1. **Comment envisagez-vous l’import de ces données ?**

Donc nous allons créer des tables en SQL en prenant les données du fichier csv avec des import. Ce qui nous permettra de faire des tables avec les données contenues dans les .csv et donc nous pouvons les modifier et faire ce que bon nous semble.

**Réponses Exercices n°2 :**

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

Nous avons créer la table **import** avec les attributs suivants :

* ID => entier
* Name => texte
* Sex => 1 caractère
* Age => entier
* Height => entier
* Weight => numérique (décimal)
* Team => texte
* NOC => texte à 3 caractères
* Games => texte à 20 caractères
* Year => entier
* Season => texte à 20 caractères
* City => texte
* Sport => texte à 40 caractères
* Event => texte
* Medal => texte à 10 caractères

1. **Remplir cette table avec les données récupérées**

Pour remplir cette table import, nous avons exécuté cette commande :

* \copy import from athlete\_events.csv with (format CSV, delimiter ',', HEADER, NULL 'NA')

1. **Certaines données sont incorrectes. C’est notamment le cas de toutes les informations avant 1920 et de toutes les épreuves “Artistiques”. Supprimez toutes ces données douteuses.**

Pour supprimer ces données, nous avons exécuté cette commande :

* DELETE FROM import WHERE year < 1920 OR Event LIKE ‘Art%’;

1. **Importer tel quel le fichier noc\_regions.csv**

Nous avons créer la table **noc** avec les attributs suivants :

* NOC => 3 caractères
* Region => texte
* Notes => texte

Puis pour importer les données nous avons fait :

* \copy noc from noc\_regions.csv with (format CSV, delimiter ',', HEADER, NULL 'NA')

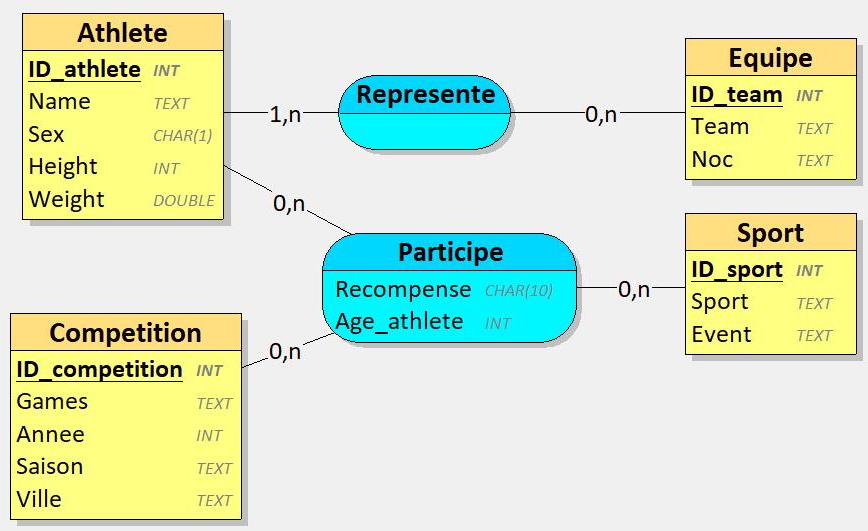
**Réponses Exercices n°4 :**

***Comment avons nous réfléchis pour la création de notre MCD :***

Tout d’abord nous nous sommes demandés quelles étaient les grandes catégories qui devaient être présentes dans notre MCD. Il était déjà évident qu’une table athlète devait être là car ils possèdent chacun des caractéristiques propres à eux. Vu que quand un athlète participe à une compétition nous somme dit qu’il était donc aussi logique d’avoir une table compétition, vu que ce sont des pays plus précisément qui y sont présents, une table équipe s'impose.

Pour plus de logique nous avons créé une table sport car un certain nombre de sports sont présents à chaque JO qui eux possèdent des caractéristiques. “Recompense” est donc dans “Participe” puisque l' athlète est récompensé seulement s’il participe à une compétition. L’âge est aussi présent car entre plusieurs compétitions, forcément il peut avoir un âge différent.

Tout ceci nous a donc conduit au mcd suivant :



Puis à partir de ce mcd nous avons pu faire ce mld :

Sport (ID\_sport, Sport, Event)

Equipe (ID\_team, Team, Noc)

Athlete (ID\_athlete, Name, Sex, Height, Weight)

Competition (ID\_competition, Ville, Games, Year, Season, ville)

Participe (#ID\_athlete, #ID\_competition, #id\_sport, recompense, age)

Represente (#ID\_athlete, #ID\_team)

Malheureusement nous sommes conscient que quand nous faisons une jointure entre toutes les tables nous n’obtenons pas 255080 lignes comme prévu mais un peu plus. Donc nous avons répondu à toutes les questions comme si nous avions bon nombre de lignes. Les requêtes renvoient les objets demandés mais certaines ont des résultats qui ne sont pas le nombre attendu.

q2.1)

Taille en octet du fichier récupéré : 41500688 octets.

q2.2)

Taille en octet de la table import : 48119808 octets.

**Réponses Exercices n°6 :**

Nous avons donc décidé de choisir un grand pays, tel que la Chine puisque nous ne connaissions pas sa réputation aux jeux olympiques et ne savions pas si ce pays faisait partie des plus forts.

Donc comme sport nous avons décidé d’observer le côté gymnastique.

Nous allons donc faire des requêtes pour voir surtout les différentes récompenses qu’à pu avoir la Chine durant ses différentes participations.

La première requête regarde justement le nombre total de médailles qu’à pu obtenir la Chine en gymnastique.

Ensuite dans une deuxième requête nous allons voir la liste de gymnastes chinois médaillés d’or qu’il y a pu avoir depuis le début des jeux olympiques.

En troisième requête nous allons donc essayer de voir le nombre de médailles gagnées par la Chine en gymnastique pour chaque année des jeux olympiques.

Puis pour finir nous allons faire la liste des événements de gymnastique où la Chine a remporté des médailles

1.Tout d’abord on va commencer par voir le nombre de médailles qu’a obtenu la Chine en gymnastique.

SELECT COUNT(\*) AS Nombre\_medailles

FROM Participe P

JOIN Sport S ON P.id\_sport = S.ID\_sport

JOIN Represente R ON P.ID\_athlete = R.ID\_athlete

JOIN Equipe E ON R.ID\_equipe = E.ID\_equipe

WHERE E.Noc = 'CHN'

AND S.Sport = 'Gymnastics'

AND P.recompense IN ('Gold', 'Silver', 'Bronze');

2.On va ensuite regarder la liste de gymnastes chinois médaillés d’or.

SELECT DISTINCT A.Name AS Nom\_gymnaste

FROM Participe P

JOIN Sport S ON P.id\_sport = S.ID\_sport

JOIN Athlete A ON P.ID\_athlete = A.ID\_athlete

JOIN Represente R ON P.ID\_athlete = R.ID\_athlete

JOIN Equipe E ON R.ID\_equipe = E.ID\_equipe

WHERE E.Noc = 'CHN'

AND S.Sport = 'Gymnastics'

AND P.recompense = 'Gold';

3.Ensuite le nombre de médailles gagnées par la Chine en gymnastique pour chaque année des jeux olympiques.

SELECT C.Year AS Annee,COUNT(\*) AS Nombre\_medailles

FROM Participe P

JOIN Sport S ON P.id\_sport = S.ID\_sport

JOIN Competition C ON P.ID\_competition = C.ID\_competition

JOIN Represente R ON P.ID\_athlete = R.ID\_athlete

JOIN Equipe E ON R.ID\_equipe = E.ID\_equipe

WHERE E.Noc = 'CHN'

AND S.Sport = 'Gymnastics'

AND P.recompense IN ('Gold', 'Silver', 'Bronze')

GROUP BY C.Year

ORDER BY C.Year ASC;

4.Pour finir la liste des événements de gymnastique où la Chine a remporté des médailles.

SELECT DISTINCT S.Event AS Evenement\_gymnastique

FROM Participe P

JOIN Sport S ON P.id\_sport = S.ID\_sport

JOIN Represente R ON P.ID\_athlete = R.ID\_athlete

JOIN Equipe E ON R.ID\_equipe = E.ID\_equipe

WHERE E.Noc = 'CHN'

AND S.Sport = 'Gymnastics'

AND P.recompense IN ('Gold', 'Silver', 'Bronze');