

## **SAE2.04 – Exploitation d'une base de donnée**

*Lelievre  
Tom*

*Thieffry  
Matteo*



Université  
de Lille

## Exercice 1 :

### Question 2 :

1. Pour savoir combien de lignes possède le fichier récupéré, nous avons fait la commande : `wc -l [nom du fichier]` . Ce qui nous renvoie 13870.  
Le fichier comporte donc 13870 lignes.
2. Une ligne représente toutes les informations liées à une formation dans un établissement durant la session ParcoursSup.
3. Pour savoir combien de colonnes possède le fichier, nous utilisons la commande : `head -n 1 data.csv | tr «;» «\n» | wc -l` . Ce qui nous renvoie 118.  
Il y a donc 118 colonnes dans le fichier.
4. La colonne qui identifie un établissement est la 3ème colonne et qui se nomme « *Code UAI de l'établissement* ».
5. Et celle qui identifie une formation est la 110ème colonne et elle se nomme « *code\_aff\_form* ».
6. Il n'y a qu'une seule ligne qui fait référence à notre BUT Informatique. Pour la trouver, nous pouvons utiliser la commande : `grep -E "BUT - Informatique" fr-esr-parcoursup.csv | grep -E "Université de Lille" | wc -l` .
7. La colonne qui identifie un département est la 5ème colonne qui se nomme « *code départemental de l'établissement* ».
8. Pour importer les données, on envisage de les importer en créant une table temporaire et en typant chaque colonne tout en ayant supprimé la première ligne.
9. On envisage plusieurs problèmes tel que par exemple dans la même case de la colonne « filière de l'établissement très agrégée car il y a du texte et en même temps des nombres entiers. Aussi, pour des colonnes tel que « Code UAI » où il y a des chiffres et des lettres.  
Et nous avons aussi des colonnes où il y a des valeurs nulles, donc si nous devons faire des calculs et qu'on tombe sur les valeurs nulles, cela nous posera un problème.

### Question 2 :

1. Pour constituer le fichier dico.xls, nous avons d'abord créé un fichier passerelle.xls pour qu'on puisse sélectionner la première ligne et la transformer en une colonne en remplaçant les «;» par des «\n». Pour faire cela, nous avons utilisé la commande : `head -1 fr-esr-parcoursup.csv | tr ";" "\n" > passerelle.xls` .

Puis avec le fichier passerelle.xls on a créé une variable qu'on a incrémenté au fur et à mesure qu'on passait une ligne tout en ajoutant «n» avant en utilisant comme commande : `awk -F ';' '{i++; printf "n%d;%s\n",i,$0}' passerelle.xls > dico.xls`

2 & 3. La table import se trouve au début du fichier parcoursup.sql

4. Pour pouvoir remplir cette table avec les données récupéré, nous ajoutons après la création de la table import : `\copy import from fr-esr-parcoursup.csv delimiter ';'`

5. Les requêtes se trouvent au début du fichier « requete.sql »

a) Au total, il y a 13869 formations qui sont gérés par ParcourSup.

b) Au total, il y a 3602 établissements qui sont gérées par ParcourSup.

c) Au total, il y a 124 formations au sein de l'Université de Lille.

d) Au total, il y a 10 formations au sein de notre IUT.

e) Le code du BUT Informatique de l'Université de Lille est : 6888 .

f) Certaines colonnes possèdent des valeurs nulles, en voici le numéro et le nom de quelques unes :

- La 36ème : Effectif des candidats classés par l'établissement en phase complémentaire

- La 47ème : Effectif total des candidats ayant accepté la proposition de l'établissement (admis)

- La 54ème : Effectif des admis néo bacheliers

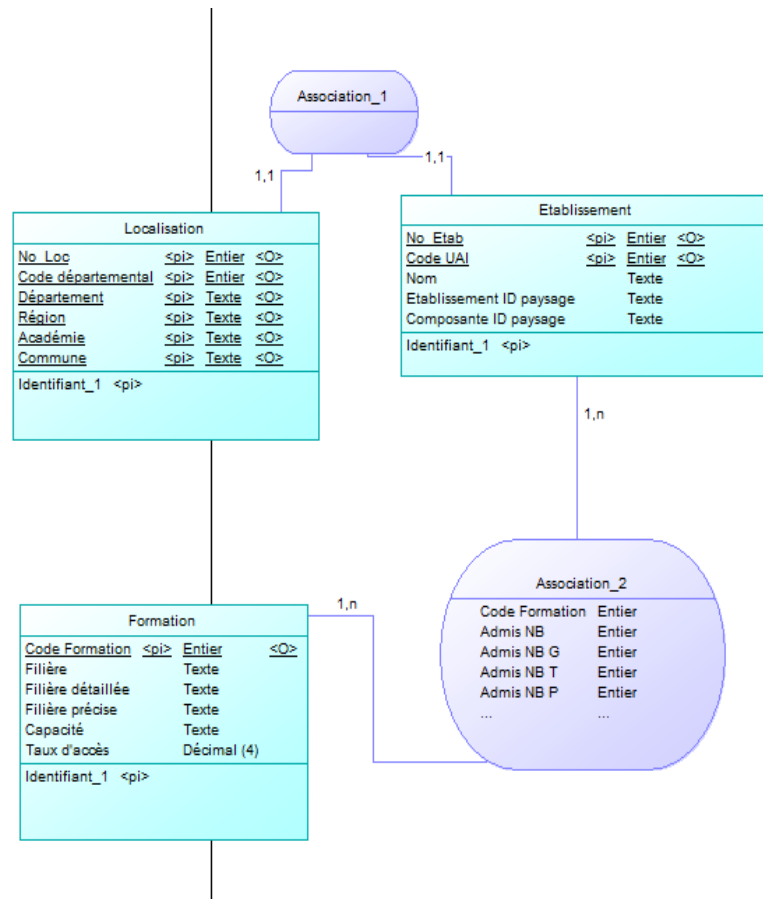
- La 70ème : % d'admis ayant reçu leur proposition d'admission à l'ouverture de la procédure principale

- La 71ème : % d'admis ayant reçu leur proposition d'admission avant le baccalauréat

Exercice 2 :

Question 1 :

1.



## Question 2 :

1. Le fichier fait 12423586 octets, ce qui est égale à 12,4 MB.
2. La table import fait un total de 13 MB soit 13000000 octets.
3. Toute les tables créées font un total de : 43KB + 152KB + 560KB + 7856KB = 8,6MB soit 8600000 octets.
4. La somme des tailles des fichiers est de 20716 + 4112 + 227727 + 6084632 = 6337187 octets, ce qui équivaut à 6,3MB.

## Exercice 3 :

Se reporter à la deuxième partie du fichier « requete.sql ».