



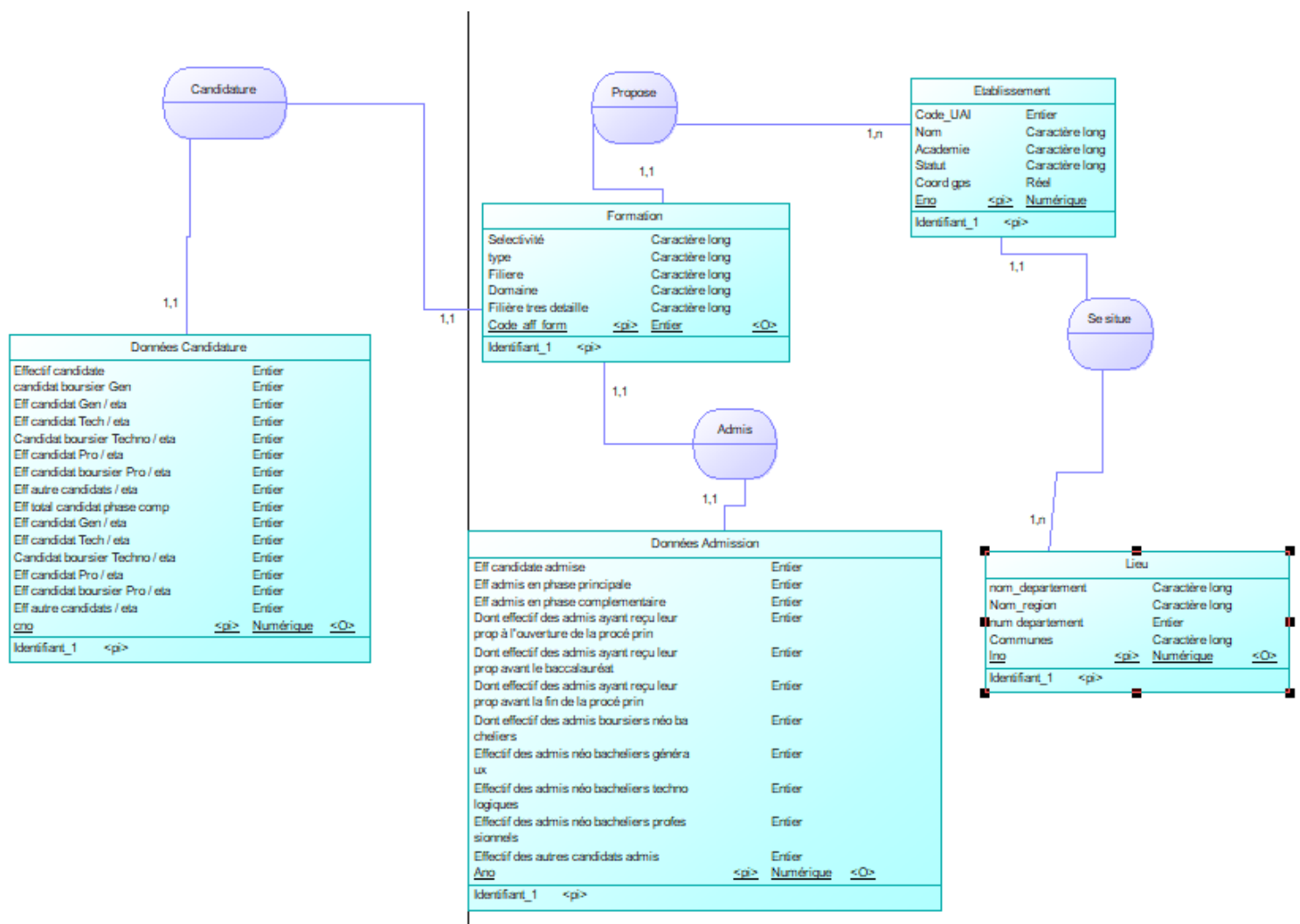
Rapport S1.04 Création de bases de données

Exploitation d'une base de données

Comment utiliser nos fichier :

- Une fois le dossier décompressé, il faut se rendre avec un terminal dans le dossier Sae2.04_Mathis_Aymane
- Ouvrir postgresql
- Exécuter le script parcoursup.sql (à l'intérieur de parcoursup.sql)
- Exécuter le script Requete.sql
- Comme demandé dans l'énoncé aucun fichier de données n'est envoyé, il est installé au début du script parcoursup.sql
- puis le fichier Données.csv est supprimé à la fin du même script *
- Si vous souhaitez conserver le fichier Données.csv il suffit de mettre la dernière ligne du script parcoursup en commentaire

Choix du MCD



Réponses au question de l'exercice 1 :

A)

Question 1 :

Il y a 13 869 lignes dans le fichier téléchargé.

Commande : `wc -l Données.csv`

(et faire -1 sur le résultat pour retirer la ligne d'en-tête)

Question 2 :

Une ligne représente une seule formation.

(par formation, nous entendons qu'il peut y avoir plusieurs BUT Informatique, mais un seul par établissement)

Question 3 :

Il y a 118 colonnes dans le fichier téléchargé.

Commande : `head -1 myfile.csv | sed 's/[,]/g' | wc -c`

(Cette commande permet de supprimer tous les caractères autres que les virgules de la ligne 1 du fichier (l'en-tête) puis elle compte le nombre de “,” ce qui permet de connaître le nombre de colonnes.

Question 4 :

La colonne UAI permet l'identification d'un établissement, c'est la 3ème colonne.

(le même identifiant peut apparaître plusieurs fois si l'établissement propose plusieurs formations.).

Question 5 :

- La colonne qui identifie une formation est la colonne "cod_aff_form", elle correspond à la 110ème colonne.

Question 6 :

Il y a une seule ligne qui fait référence à notre BUT Informatique.

Commande : **select n110 from import where n4 like '%Institut universitaire de technologie de Lille%' and n10 like 'BUT - Informatique';**

Renvoie 6888 qui est le code associé au BUT Info de l'IUT A

Question 7 :

- La colonne qui identifie le département et la colonne 5 "Code départemental de l'établissement" ou 6 "Département de l'établissement"

Question 8 :

Nous envisageons d'utiliser une table import contenant toutes les lignes du tableau avec pour type le type "Text". (le type Text permet de s'assurer qu'il n'y ait aucun problème lors de l'import avec les types de données.).

Question 9 :

- Certaines lignes ont des colonnes null ce qui peut poser problème dans les futures tables si elles ne sont pas bien créées.
- Certaines colonnes ont exactement le même nom. (filière de formations)

- Plusieurs colonnes sont inutiles dans notre cas (comme les statistiques sur les demandeurs d'internat)
- Et pour finir, le fichier téléchargé contient des lignes "total" ou "%" qui sont calculables et qui doivent être supprimés de la base de données.

B)

Question 1 :

Pour faire correspondre les noms avec les numéros dans le dico nous avons d'abord sélectionné les noms des colonnes puis copié-collé dans dico.xls, ensuite dans la ligne d'en dessous nous avons entré n1 , n2 puis nous avons glissé jusqu'à la fin des colonnes non-vides (fichier dico.xls dans le dossier compressé)

Question 2 :

- Commande SQL :
Create table import(
n1 text, n2 text, n3 text, n4 text, n5 text, ..., n118 text
)

Question 4 :

```
\copy import from ../Data/Données.csv delimiter
',' csv header;
```

Question 5a :

- Il y a 13 869 formations gérées par Parcoursup.
Commande : select count() from import ;

Question 5b :

- Il y a 3965 établissements gérés par ParcourSup.
Commande : select count(distinct n3) from import ;

Question 5c :

- Il y a 146 formations pour l'université de Lille.
Commande : select count(*) from import where n4 like '%Université de Lille%' ;

Question 5d :

- Il y a 10 formations pour notre IUT.
Commande : select count(*) from import where n4 like '%Institut universitaire de technologie de Lille%';

Question 5e :

- Le code du BUT Informatique de l'université de Lille est 6888.
Commande : **select n110 from import where n4 like '%Institut universitaire de technologie de Lille%' and n10 like 'BUT - Informatique';**

Question 5f :

- 5 colonnes contenant des cases nulle peuvent être :
 - la colonne n°9
 - la colonne n°17
 - la colonne n°22
 - la colonne n°25
 - la colonne n°26

Réponses au question de l'exercice 2 :

Question 1 :

- (Une capture d'écran du MCD se trouve en page 2 de ce rapport)

Question 2a :

- La taille que fait le fichier récupéré est de 12 416 245 octets soit 12.4 Mo.
- **commande** : `wc -c ../Data/Données.csv`

Question 2b :

- La taille en octet de la table import est de 16 228 352 octets soit 16.2 Mo
- **Requêtes** : `select
pg_total_relation_size('import');`

Question 2c :

- La taille en octet que fait la somme des tables créées est de 7 585 792 octets soit 7,58 Mo.
- ```
select pg_total_relation_size('Etablissement')+
pg_total_relation_size('Formation')+
pg_total_relation_size('Lieu')+
pg_total_relation_size('Données_admission')+
pg_total_relation_size('Donnees_candidature')
as taille_somme_table;
```

### Question 2d :

- La taille en octet que fait la somme des tailles des fichiers exportés correspondant à ces tables est de 3 836 637 octets soit 3.83 Mo.
- ```
wc -c ../tables_exportée/Lieu.csv -c  
../tables_exportée/Formation.csv -c  
../tables_exportée/Donnees_candidature.csv  
-c ../tables_exportée/Données_admission.csv  
-c ../tables_exportée/Etablissement.csv
```


Réponses au question de l'exercice 3 :

(des commentaires ont été placé à l'aide de la commande echo dans le script)

Question 1 :

- Requête : `select n56,(n57+n58+n59) as calcul from import group BY n56,n57,n58,n59 ;`

Question 2 :

- Requête : `select n56-(n57+n58+n59) as verification from import where n56-(n57+n58+n59) <> 0 group by n56,n57,n58,n59;`

Question 3 :

- Requête : `select n74,round(((n51*100)/n47)) as calcul from import where n47 <> 0 group by n74,n51,n47;`

Question 4 :

- Requête : `select n74-round(((n51*100)/n47)) as calcul from import where n47<> 0 and (n74-round(((n51*100)/n47))) <>0 group by n74,n51,n47;`

Question 5 :

- Requête : `select n76,round(((n53*100)/n47)) as calcul from import where n47 <> 0 group by n76,n53,n47;`

Question 6 :

- Requête :

```
select
round((Admis_avant_fin_procedure_princ*100)/Total_accept_
prop)
from Données_admission
where Total_accept_prop <>0
group by
Admis_avant_fin_procedure_princ,Total_accept_prop
limit 10;
```

Question 7 :

- Requête :

```
select n81, round(round(n55)/round(n56)*100) as
calcul from import where n56 <> 0 group by n81,n55,n56;
```

Question 8 :

- Requête :

```
select
round(round(Admis_boursiers_neo_bac)/round(Admis_neo_bac
techno+Admis_neo_bac_gen+Admis_neo_bac_pro)*100) as
calcul from Données_admission
where
round(Admis_neo_bac_techno+Admis_neo_bac_gen+Admis_neo_ba
c_pro)<>0
group by
Admis_boursiers_neo_bac,Admis_neo_bac_gen,Admis_neo_bac_p
ro,Admis_neo_bac_techno order by calcul des
limit 10 ;
```