

# SAÉ 2.04

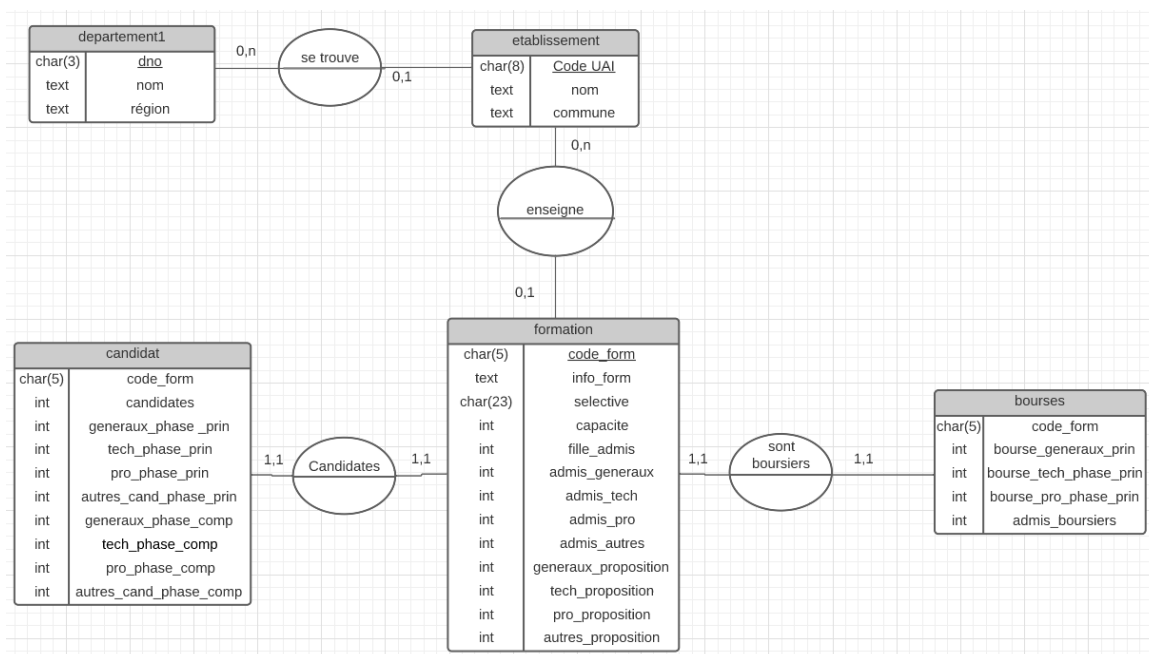
## Exploitation d'une BD

Cédric Larsonnier

Milan Delzenne

05-04-2022

MCD :



## Introduction :

Lors de cette SAE 2.4 de la partie BD nous avons eu une base de données qui répertorie toutes les inscriptions des étudiants à Parcoursup. Le but était de comprendre un ensemble de données réelles, savoir les importer, les formaliser et réaliser des requêtes classiques.

## Compréhension des données :

Au début nous devons comprendre les données de Parcoursup :

- Grâce à la commande “wc -l fr-esr-parcoursup.csv” cela vas nous donne 13397 ligne
- Chaque ligne correspond à une filière des établissements.
- Grâce à la commande “ head -n 1 fr-esr-parcoursup.csv | tr -s ‘,’ ‘\n’ | wc -l ” cela va nous donner 118 colonnes.

## Entête de chaque colonne numéroté de 1 à 118 :

1. Session;	65. Dont effectif des admis néo bacheliers avec mention Très Bien avec félicitations au bac;
2. Statut de l'établissement de la filière de formation (public, privé...);	66. Effectif des admis néo bacheliers généraux ayant eu une mention au bac;
3. Code UAI de l'établissement;	67. Effectif des admis néo bacheliers technologiques ayant eu une mention au bac;
4. Établissement;	68. Effectif des admis néo bacheliers professionnels ayant eu une mention au bac;
5. Code départemental de l'établissement;	69. Dont effectif des admis issus du même établissement (BTS/CPGE);
6. Département de l'établissement;	70. Dont effectif des admises issues du même établissement (BTS/CPGE);
7. Région de l'établissement	71. Dont effectif des admis issus de la même académie;
8. ;Académie de l'établissement;	72. Dont effectif des admis issus de la même académie (Paris/Créteil/Versailles réunies);
9. Commune de l'établissement;	73. % d'admis ayant reçu leur proposition d'admission à l'ouverture de la procédure principale;
10. Sélectivité;	74. % d'admis ayant reçu leur proposition d'admission avant le baccalauréat;
11. Filière de formation très agrégée;	75. % d'admis ayant reçu leur proposition d'admission avant la fin de la procédure principale;
12. Filière de formation détaillée	76. % d'admis dont filles;
13. ;Filière de formation;	77. % d'admis néo bacheliers issus de la même académie;
14. Filière de formation détaillée bis;	78. % d'admis néo bacheliers issus de la même académie (Paris/Créteil/Versailles réunies);
15. Filière de formation très détaillée;	79. % d'admis néo bacheliers issus du même établissement (BTS/CPGE);
16. Coordonnées GPS de la formation;	80. % d'admis néo bacheliers boursiers;
17. Capacité de l'établissement par formation;	81. % d'admis néo bacheliers;
18. Effectif total des candidats pour une formation;	82. % d'admis néo bacheliers sans information sur la mention au bac;
19. Dont effectif des candidates pour une formation;	83. % d'admis néo bacheliers sans mention au bac;
20. Effectif total des candidats en phase principale;	84. % d'admis néo bacheliers avec mention Assez Bien au bac;
21. Dont effectif des candidats ayant postulé en internat;	85. % d'admis néo bacheliers avec mention Bien au bac;
22. Effectif des candidats néo bacheliers généraux en phase principale;	86. % d'admis néo bacheliers avec mention Très Bien au bac;
23. Dont effectif des candidats boursiers néo bacheliers généraux en phase principale;	87. % d'admis néo bacheliers avec mention Très Bien avec félicitations au bac;
24. Effectif des candidats néo bacheliers technologiques en phase principale;	88. % d'admis néo bacheliers généraux;
25. Dont effectif des candidats boursiers néo bacheliers technologiques en phase principale;	89. Dont % d'admis avec mention (BG);
26. Effectif des candidats néo bacheliers professionnels en phase principale;	90. % d'admis néo bacheliers technologiques;
27. Dont effectif des candidats boursiers néo bacheliers professionnels en phase principale;	
28. Effectif des autres candidats en phase principale;	
29. Effectif total des candidats en phase complémentaire;	
30. Effectif des candidats néo bacheliers généraux en phase complémentaire;	
31. Effectif des candidats néo bacheliers technologique en phase complémentaire;	
32. Effectif des candidats néo bacheliers professionnels en phase complémentaire;	

33. Effectifs des autres candidats en phase complémentaire;	91. Dont % d'admis avec mention (BT);
34. Effectif total des candidats classés par l'établissement en phase principale;	92. % d'admis néo bacheliers professionnels;
35. Effectif des candidats classés par l'établissement en phase complémentaire;	93. Dont % d'admis avec mention (BP);
36. Effectif des candidats classés par l'établissement en internat (CPGE);	94. Effectif des candidats en terminale générale ayant reçu une proposition d'admission de la part de l'établissement;
37. Effectif des candidats classés par l'établissement hors internat (CPGE);	95. Dont effectif des candidats boursiers en terminale générale ayant reçu une proposition d'admission de la part de l'établissement;
38. Effectif des candidats néo bacheliers généraux classés par l'établissement;	96. Effectif des candidats en terminale technologique ayant reçu une proposition d'admission de la part de l'établissement;
39. Dont effectif des candidats boursiers néo bacheliers généraux classés par l'établissement;	97. Dont effectif des candidats boursiers en terminale technologique ayant reçu une proposition d'admission de la part de l'établissement;
40. Effectif des candidats néo bacheliers technologiques classés par l'établissement;	98. Effectif des candidats en terminale professionnelle ayant reçu une proposition d'admission de la part de l'établissement;
41. Dont effectif des candidats boursiers néo bacheliers technologiques classés par l'établissement	99. Dont effectif des candidats boursiers en terminale générale professionnelle ayant reçu une proposition d'admission de la part de l'établissement;
42. ;Effectif des candidats néo bacheliers professionnels classés par l'établissement;	100. Effectif des autres candidats ayant reçu une proposition d'admission de la part de l'établissement;
43. Dont effectif des candidats boursiers néo bacheliers professionnels classés par l'établissement;	101. Regroupement 1 effectué par les formations pour les classements;
44. Effectif des autres candidats classés par l'établissement;	102. Rang du dernier appelé du groupe 1;
45. Effectif total des candidats ayant reçu une proposition d'admission de la part de l'établissement;	103. Regroupement 2 effectué par les formations pour les classements;
46. Effectif total des candidats ayant accepté la proposition de l'établissement (admis);	104. Rang du dernier appelé du groupe 2;
47. Dont effectif des candidates admises;	105. Regroupement 3 effectué par les formations pour les classements;
48. Effectif des admis en phase principale;	106. Rang du dernier appelé du groupe 3;
49. Effectif des admis en phase complémentaire;	107. list_com;
50. Dont effectif des admis ayant reçu leur proposition d'admission à l'ouverture de la procédure principale;	108. Taux d'accès des candidats ayant postulé à la formation (ratio entre le dernier appelé et le nombre vœux PP);
51. Dont effectif des admis ayant reçu leur proposition d'admission avant le baccalauréat;	109. tri;
52. Dont effectif des admis ayant reçu leur proposition d'admission avant la fin de la procédure principale;	110. COD_AFF_FORM;
53. Dont effectif des admis en internat;	111. LIB_FOR_VOE_INS;
54. Dont effectif des admis boursiers néo bacheliers;	112. detail_forma2;
55. Effectif des admis néo bacheliers;	113. Lien de la formation sur la plateforme Parcoursup;
56. Effectif des admis néo bacheliers généraux;	114. Dont taux d'accès des candidats ayant un bac général ayant postulé à la formation;
57. Effectif des admis néo bacheliers technologiques;	115. Dont taux d'accès des candidats ayant un bac technologique ayant postulé à la formation;
58. Effectif des admis néo bacheliers professionnels;	116. Dont taux d'accès des candidats ayant un bac professionnel ayant postulé à la formation;
59. Effectif des autres candidats admis;	117. etablisement_id_paysage;
60. Dont effectif des admis néo bacheliers sans information sur la mention au bac; Q4	118. composante_id_paysage;
61. Dont effectif des admis néo bacheliers sans mention au bac;	
62. Dont effectif des admis néo bacheliers avec mention Assez Bien au bac;	
63. Dont effectif des admis néo bacheliers avec mention Bien au bac;	
64. Dont effectif des admis néo bacheliers avec mention Très Bien au bac;	

### **Information sur certaine colonne importante :**

Sur le fichier csv de Parcoursup il y a plusieurs colonne à prendre en compte :

- colonne C le Code UAI des établissement
- colonne D pour le nom de l'établissement
- colonne M pour le nom des formations
- colonne DF pour le code de formations
- colonne E le code départementale de l'établissement
- colonne F nom du département de l'établissement

### **Information complémentaire :**

Il y a 10 lignes qui référencent notre établissement (I.U.T. de Lille A) les lignes 681,682,4956,6223,8988,9709,9710,9711,10829,10830.

Pour importer toutes les données nous allons utiliser une table import afin de les mettre les données dans d'autre table créée par la suite.

Le gros problème des fichiers csv est la recherche de données, trouver des information nécessaire rapidement est quasi impossible.

### **Importation des table dans postgres :**

Pour importer les tables nous avons besoin d'enlever les entêtes qui est inutile, grâce à cette commande les première ligne du fichier csv sont enlever :

```
“\copy import from fr-esr-parcoursup.csv delimiter ';' ”
```

Comme le fichier de Parcoursup a plus de 118 colonnes il faut ventiler les données afin d'avoir un MCD lisible et avoir la possibilité de l'intégrer dans postgres facilement, alors nous avons créé 5 tables différentes :

- departement1
- etablissement
- formation
- candidat
- bourses

Chacune de ces table a des colonne bien spécifique liée au données de la table import.

### **Requête sur la table import :**

Cette requête nous permet de s'avoir combien de formation il y a dans Parcoursup :

```
SELECT COUNT(DISTINCT n109) FROM import;  
-13396
```

Cette requête nous dit combien de combien d'établissement son gère par Parcoursup :

```
SELECT COUNT(DISTINCT n2) FROM import;  
-3869
```

Cette requête nous informe le nombre de formation propose l'Université de Lille :

```
SELECT COUNT(*) FROM import WHERE n3 = 'Université de Lille';  
-118
```

Cette requête nous informe du nombre de formations proposé par L'I.U.T. A  
SELECT COUNT(n3) FROM import WHERE n3 LIKE 'I.U.T. de Lille A%';  
-10

Ces requete montre des colonnes qui ont des valeur null :  
SELECT n14, n20, n35, n36, n52 FROM import;

### **Taille des fichiers :**

Le fichier csv récupéré est actuellement à 12,7 Mo.

Cette commande donne le nombre d'octets de la table import :  
SELECT pg\_size\_pretty( pg\_total\_relation\_size('import'));  
La table import fait 15 Mo

Les tables créer :

-departement1 : 32 kB	-candidat : 1032 kB
-etablissement : 488 kB	-bourses : 712 kB
-formation : 2936 kB	

-TOTAL : 5,203 Mo

La taille total des table exporté font 2,652 kB

### **Utilisation des tables créer :**

**-Cette commande affiche la colonne n55 (effectif des admis néo bachelier) et une colonne qui le recalcule :**

SELECT n55, n56 + n57 + n58 FROM import;

**Vérifie que le calcul soit exact :**

SELECT COUNT(\*) FROM import WHERE n55 = n56+n57+n58;

**-Cette commande affiche la colonne n73 (% des d'admis ayant reçu leur proposition d'admission à l'ouverture de la procédure principale) et une colonne qui le recalcule :**

```
SELECT n73, ROUND(CAST(n50 as numeric)/n46*100.0,2) FROM import WHERE n46 <> 0;
```

**Vérifie que le calcul soit exact :**

```
SELECT COUNT(*) FROM import WHERE CAST(n73 as numeric) =  
ROUND(CAST(n50 as numeric)/n46*100.0,2) AND n46<>0;
```

**-Cette commande affiche la colonne n76 (% d'admis dont filles) et une colonne qui le recalcule à 1 décimale près :**

```
SELECT n76, (100.0*n47)/n46 FROM import WHERE n46 <> 0;
```

**Vérifie sur les tables ventilé que la colonne n76 peut se calculer :**

```
SELECT  
(100.0*filles_admis)/(admis_generaux+admis_tech+admis_pro+admis_autres) FROM  
formation WHERE (admis_generaux+admis_tech+admis_pro+admis_autres) <>0;
```

**-Cette commande affiche la colonne n81 (% d'admis néo bacheliers) et une colonne qui le recalcule à 1 décimale près :**

```
SELECT n81, n55*100.0/n46 FROM import WHERE n46 <> 0;
```

**Vérifie sur les tables ventilé que la colonne n81 peut se calculer :**

```
SELECT  
(admis_generaux+admis_tech+admis_pro)*100.0/(admis_generaux+admis_tech+admi  
s_pro+admis_autres) FROM formation WHERE  
(admis_generaux+admis_tech+admis_pro+admis_autres) <> 0;
```