

Partie 1 - Implantation du schéma de la base de données

Introduction au projet

L'objectif de ce projet est vous faire implanter, peupler puis exploiter une base de données réaliste. Nous allons nous appuyer sur vos compétences de traduction de modèles de données (diagramme de classe UML) et de définition d'un schéma relationnel en SQL, mais aussi de Mathématiques pour exploiter des statistiques tirées de cette base de données, de gestion de projet pour la planification, et d'Anglais pour la présentation de votre travail.

Le sujet de cette année porte sur les admissions avec **ParcourSup** dans toutes les formations de France (13644 dans notre fichier de 2022). Pour cela nous allons vous fournir dans un premier temps un diagramme de classe à traduire dans le modèle relationnel. Dans une prochaine partie, nous vous fournirons des données présentées sous forme d'un fichier CSV qui ont été extraites de l'application **ParcourSup**. Ces données son en libre accès sur internet. Pour simplifier, nous ne traiterons que les données de 2022, même si nous représenterons la notion de session (d'année) dans le diagramme de classe. En effet, il est très probable que les fichiers changent de format d'une année sur l'autre.

Présentation du modèle de données

Comme indiqué dans l'introduction, nous souhaitons modéliser une partie des données nécessaires à l'étude des candidatures et des admissions dans **ParcourSup**.

Pour cela nous vous proposons le diagramme de classe dans la figure 1.

Le diagramme de classe représente les informations concernant les différentes formations gérées dans **ParcourSup**. Ces formations sont dispensées dans un établissement, sur un site géographique (commune dans un département et dans une région), et dépend d'une académie du ministère de l'Education Nationale. Chacune de ces formations correspond à une filière. Par exemple, la filière du BUT Informatique peut être située à Lannion, mais aussi à Vannes, dépendre de l'académie de Rennes, de l'Université de Rennes ou de l'université de Bretagne Sud. Le BUT informatique de l'IUT de Lannion est donc une formation.

Ensuite, les données concernant les candidatures à ces formations et les admissions qui ont suivi sont enregistrées. Chacune des valeurs dépend donc au moins de la formation et de la session (année) ¹. Il y des généralités sur les admissions (sélectivité, capacité, nombre total d'admis et d'admisses, *etc.*). Il y a aussi des informations selon les types de baccalauréats des candidats et des admis, selon la mention obtenu au baccalauréat, et enfin des informations sur le rang du dernier appelé selon un critère de regroupement choisi par la formation (Tous les bacs généraux, Tous les boursiers, *etc.*).

Certains attributs auront des valeurs fixées qui ne sont pas données de façon évidente dans le CSV. Voici les valeurs pour le type de bac (**type_bac**) et pour la mention (**libelle_mention**).

type_bac
Bac général
Bac technologique
Bac professionnel
Autres

libelle_mention
Sans information
Sans mention
Assez bien
Bien
Très bien
Très bien avec félicitations du jury

1. Même si nous ne manipulerons que les données de 2022 dans un premier temps, plus si nous en avons le temps.

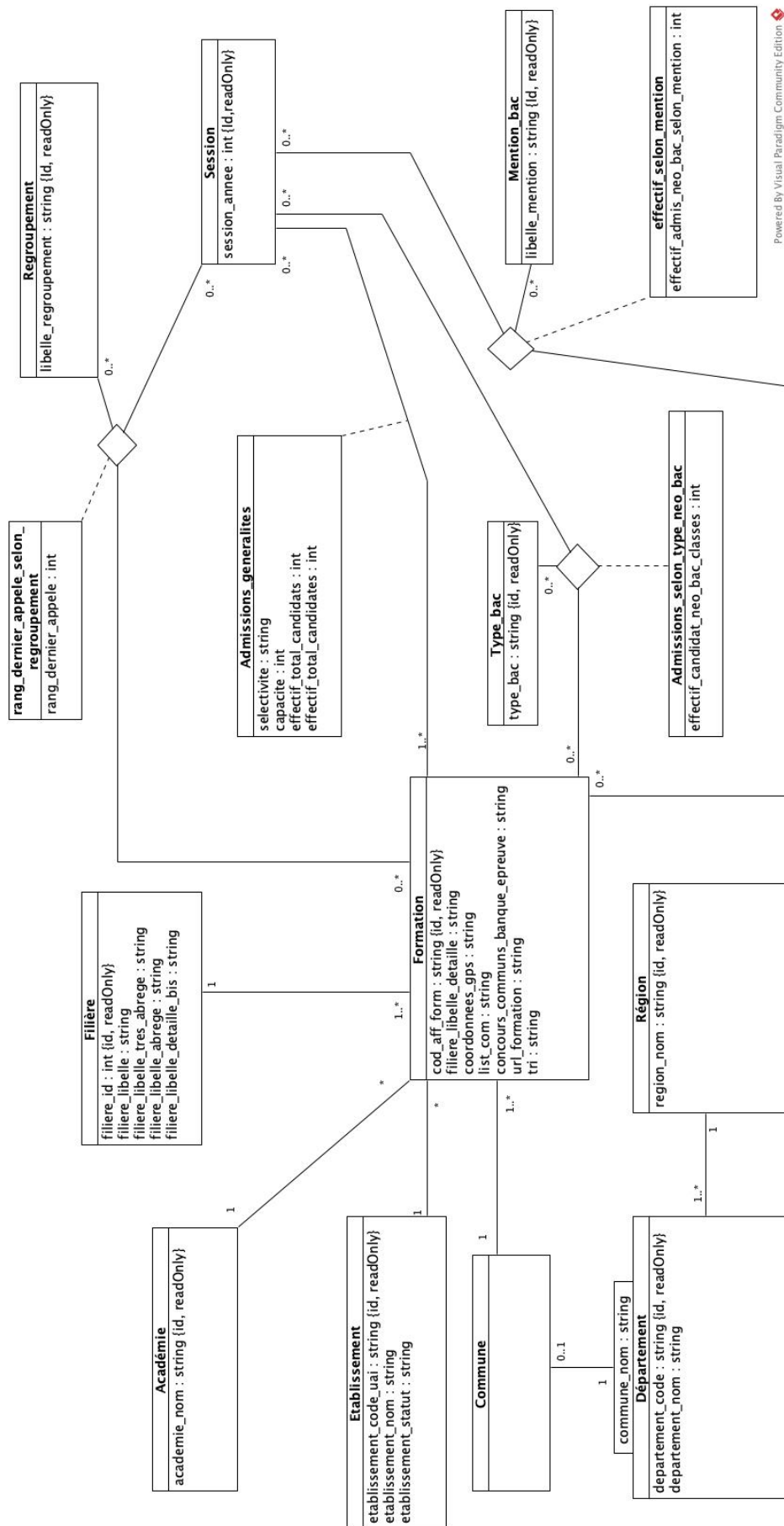


FIGURE 1 – Gestion des candidatures et des admissions dans ParcoursUp - Diagramme de classe

Ce dont vous disposez pour travailler

- Le présent énoncé `sae204_2023_partie1.pdf`.
- un fichier tableur nommé `dico_data_toute_annee.xlsx` décrivant le dictionnaire des données pour vous aider à la compréhension du diagramme. Des couleurs sont utilisées pour vous indiquer quelles propriétés ont été gardées dans le diagramme. Ont été écartées les données calculables qui ne doivent évidemment pas être stockées. Vous pourrez réutiliser les noms de la première colonne qui sont ceux utilisés dans le diagramme de classe (à quelques détails près).
- un fichier texte `proprietes.txt` contenant toutes les propriétés du CSV et celles utilisées dans le diagramme de classe. Il pourra vous servir pour faire des copier-coller pour votre schéma relationnel en SQL.
- Un fichier CSV contenant toutes les données évoquées ci-avant et dans le diagramme de classe, mais bien plus encore car nous avons choisi de limiter le projet. Ce fichier se nomme `fr-esr-parcoursup_2022.csv`. Vous pourrez constater que les entêtes de colonne ne sont pas exactement les mêmes que les noms de propriétés choisis dans le diagramme de classe. Par contre vous retrouverez le nom des propriétés dans le dictionnaire des données exactement dans le même ordre que dans le fichier CSV.
- le diagramme de classe sous forme d'un fichier image sous le nom de `ParcourSup_simple.jpg`

Mise en œuvre et travail demandé

- **Réfléchir au schéma relationnel** : graphe des contraintes d'intégrité référentielles, traduction du diagramme de classes sous forme d'un script SQL de création de bout en bout de la structure de la base de données. Ce script sera à déposer sur Moodle dans un dépôt dédié créé pour l'occasion.
- **Implantation sur votre base** commune nommée `pg_sae204_k12` par exemple, si vous êtes le deuxième binôme du groupe K1. Pour cela il vous faut constituer des binômes sur [ce fichier Excel qui est en ligne](#). Cette implantation consiste à traduire le diagramme de classe en un schéma relationnel exprimé en SQL.

Les consignes suivantes devront **obligatoirement être respectées** :

- L'implantation du schéma relationnel de la base de données sera réalisée avec le SGBD PostgreSQL à l'aide d'un script SQL que vous nommerez `parcoursup_db_create.sql`. Ce script est supposé pour s'exécuter de bout en bout et réinitialiser le schéma et définir sa structure.
- Les noms des schémas seront définis en minuscules.
- Les relations (tables) seront définies au sein d'un schéma nommé `parcoursup`.
- Les noms des relations de base seront définis en minuscules et préfixés par `_`. Exemples : `_departement`, `_formation`.
- Les contraintes seront nommées explicitement.
- les noms d'attributs seront définis en minuscules et reproduiront les variables d'instance ou noms de rôle (au singulier) du diagramme de classes. Etant donné le nombre important de propriétés dans le diagramme de classe et encore plus dans le fichier CSV, pour vous aider, nous vous fournissons une annexe qui contient une liste des attributs sous la forme d'un dictionnaire des données.

Livrables attendus pour cette première partie (date de rendu : 19/05/2023 à 18h00 au plus tard) :

- un fichier `parcoursup_db_create.sql` à déposer sur l'ENT
- l'implantation de la structure de la base à l'aide de ce script dans un schéma nommé `parcoursup` sur votre base commune `pg_sae204_k12` (si vous êtes membre du groupe K12).