
Partie 2 - Peuplement de la base de données

Description des données fournies

L'objectif de cette partie 2 est de vous faire peupler la base modélisée dans la partie 1. Vous disposez pour cela de plusieurs fichiers de données qui sont extraites de différentes applications utilisées à l'IUT pour gérer les admissions, les inscriptions et la scolarité et enfin les résultats. Vous disposez également du schéma de la base de données dans un fichier nommé `sae204_partie2_schema.sql`. Le schéma a été simplifié sur les identifiants pour vous faciliter le peuplement de la base.

Les fichiers sont les suivants :

- `v_candidatures.csv` : données liées aux candidatures (APB puis Parcoursup) . Il y a bien sûr plus de candidats que d'inscrits.
- `v_inscriptions.csv` : données liées aux inscriptions à l'IUT.
- `v_resu_s1.csv`, `v_resu_s2.csv`, `v_resu_s3.csv` et `v_resu_s4.csv` : les résultats des étudiants inscrits.
- `ppn.csv` : la description des différents modules par semestre.
- `v_programme.csv` : les modules prévus au programme lors des différentes années universitaires.

Comme les applications sont différentes, les données entre ces différents *référentiels* sont saisies lors de chaque étape de leur utilisation. Entre chaque référentiel, il peut y avoir des incohérences, des données manquantes, de la redondance d'information que vous retrouverez dans les extractions fournies. Votre objectif est de corriger ces défauts au maximum lors de l'importation de ces données dans votre base de données, qui est unique, elle. Vous pourrez concevoir des requêtes pour vérifier la cohérence de ces données (ex : les étudiants ayant des résultats sont-ils tous inscrits?).

Quelques remarques d'aide à l'exploitation des données

Vous pourrez constater que certains champs dans les fichiers CSV sont vides. Vous aviez peut-être prévu des NOT NULL pour tous vos attributs dans votre base, ce qui est un bon réflexe. Vous allez maintenant devoir relâcher un peu la contrainte dans le schéma relationnel qui vous est fourni pour pouvoir importer les données. Vous devrez proposer des commandes pour relâcher ces contraintes, en partant du schéma décrit dans le fichier `sae204_partie2_schema.sql`.

Les résultats sont présentés sous forme d'un n-uplet pour un étudiant et un semestre d'une année donnée. Il va falloir décomposer ces lignes car vous aurez certainement prévu un n-uplet pour un étudiant et un module dans un semestre d'une année donnée.

Vous verrez que toutes les notes ne sont pas des nombres. Il existe deux valeurs spéciales : `~` (tilde) qui marque l'absence de note, et `-c-` qui marque une note capitalisée parce qu'une UE a été validée lors de la tentative précédente sur le semestre, lors d'un redoublement. Une moyenne générale de 10 ou plus, une moyenne d'UE à 10 ou plus permet de capitaliser cette UE entière pour un redoublement. La règle d'obtention d'un semestre est d'obtenir 8 dans chaque UE et 10 de moyenne générale.

Il y a sûrement d'autres questions que vous pourrez vous poser sur ces données. Bonne exploration et bon courage dans le monde cruel des référentiels multiples !

Mise en œuvre et travail demandé

- **Peupler la base** : utilisez les fichiers CSV et la commande `WbImport` et tout moyen technique que vous trouverez utile pour importer vos données dans les tables que vous avez définies.
- A propos de ce peuplement, vous aurez une réflexion sur l'efficacité computationnelle et par conséquent énergétique de votre procédure. Expliquez, dans un paragraphe commenté en début de script, l'idée générale de votre méthode et pourquoi vous pensez qu'elle est efficace.
- Optionnel : **Imaginer des vues, sans les implanter pour l'instant** pour faciliter la visualisation des données et le calcul de statistiques : calcul de moyennes par ue, par semestre, à partir de ces moyennes, décisions suggérées pour la CPJ¹. Cela vous servira pour la partie 3 de la SAE.

1. Commission Préparatoire au Jury

Les consignes suivantes devront obligatoirement être respectées :

- Le peuplement du schéma relationnel de la base de données sera réalisée avec le SGBD PostgreSQL à l'aide d'un script SQL que vous nommerez `sae204_xyz_partie2.sql`, si `xyz` est votre identifiant de binôme. Ce script part de la situation décrite dans le fichier SQL qui vous est fourni. Vous déposerez votre propre script sur l'ENT **avant le vendredi 13 juin à 18h00**.
- Les relations (tables) seront définies au sein d'un schéma nommé `partie2`. Les données devront être présentes dans le même schéma à la date limite de rendu.
- Votre script s'exécutera de bout en bout, sans interruption, pour modifier les contraintes, le cas échéant, et peupler la base.
- Le résultat de votre script devra être présent sur le serveur `servbdd` de l'IUT au moment de votre rendu (date limite le **vendredi 13 juin 2025 à 18h00** sur l'ENT).