
Partie 2 - Peuplement 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 admission, les inscriptions et la scolarité et enfin les résultats.

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.

Comme les applications sont différentes, les données entre ces différents *référentiels* sont saisies lors de chaque étape de leur utilisation. Entre chaque référentiels, 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 vérifie la cohérences 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 des champs dans les fichiers CSV sont vides. Vous avez peut-être prévu des NOT NUL pour tous vos attributs dans votre base, ce qui est un bon réflexe. Vous allez maintenant devoir relâcher un peu la contrainte pour pour voir importer les données.

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 règles 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 (tables temporaires?) pour importer vos données dans les tables que vous avez définies.
- **Imaginer des vues** 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 CPJ ¹.

1. Commission Préparatoire au Jury