

Mini-projet : Conception et développement d'une base de données - PARTIE 1

Présentation

Ce mini-projet reprend toutes les étapes de la méthode MERISE vues dans ce module et utilisées dans l'industrie pour la conception et le développement d'une base de données. A partir d'un domaine qui vous intéresse, vous allez construire votre base de données en suivant les étapes suivantes :

Etape 1 : Analyser les besoins

- Choisir un domaine qui vous intéresse
- Utiliser l'IAG pour obtenir des informations sur les règles métier de votre domaine, et les données à stocker

Etape 2 : MCD

- A partir des données obtenues à l'étape 1, produire le MCD
- Il doit être normalisé et inclure des éléments avancés de modélisation

Etape 3 : MLD et MLP

- A partir du MCD, récupérer le MLD et MPD en SQL
- Définir les contraintes additionnelles et le code SQL correspondant

Etape 4 : Insérer des données

- Remplir la base de données en s'aidant de l'IAG
- S'assurer du respect des contraintes d'intégrité des données générées

Etape 5 : Interroger la base

- Imaginer un scénario d'utilisation de votre BD
- Ecrire les requêtes SQL DML correspondant à ce scénario

Etape 6 : Présentation

- Présentation de votre travail à travers une vidéo
- Explication des choix de modélisation et d'utilisation de la BD

- La **partie 1** du projet correspond aux **étapes 1 et 2**, la **partie 2** aux **étapes 3, 4, 5 et 6**.
- Ce projet compte pour **20%** de la note du module.
- Le projet est à réaliser en **binôme**. Un seul trinôme est autorisé en cas de nombre d'étudiants impair.

- Il est divisé en deux parties :
 - Le sujet de la 1ère partie du projet est diffusé le **13/02/26** et elle doit être rendue le **27/02/26**,
 - Le sujet de la 2ème partie est diffusé le **27/02/26** et elle doit être rendue pour le **16/03/26**.
- Vous devrez rendre (*liste non-exhaustive, le détail vous est donné dans la suite*) :
 - un fichier de documentation **README.md**,
 - différents **fichiers texte** qui ont servi à la conception et insertion des données (prompts),
 - des **fichier exécutables .sql** ,
 - une **vidéo** de présentation.
- L'ensemble des fichiers doit être stocké dans un **répertoire Github** partagé avec votre enseignant de TP. Il n'y a rien à lui envoyer, *chaque rendu est juste un dépôt sur le répertoire Github*.

La **partie 1** correspond à l'analyse des besoins et la création du MCD.

I. Première étape : analyse des besoins

I.A. Récupérer les besoins

L'analyse des besoins correspond à obtenir toutes les informations nécessaires sur ce qui doit être stocké. Ces informations correspondent à un sujet, un domaine particulier. Dans le monde professionnel, elles sont obtenus en communiquant avec l'entreprise qui veut créer sa base de données.

En exercices, vous avez utilisé des descriptions qui incluait des règles métier. Ex : "Un employé ne peut pas être affilié à plus d'un service", "Les employés ont un manager" ou bien "Un service a un nom qui a la forme "S" + numéro sur 2 digits" sont des règles métier.

Pour ce projet, votre interlocuteur ne sera pas une entreprise/un organisme/une administration, mais une IAG qui tiendra le rôle d'une entreprise/un organisme/une administration.

I.B. Prompt engineering

Pour interroger l'IAG, vous allez utiliser un prompt qui utilise un cadre approprié : le framework RICARDO, qui permet d'écrire de "bons" prompts.

Lettre	Signification	Détail
R	Rôle	Définit le rôle de l'IA générative (<i>par exemple, « Agissez en tant qu'enseignant » ou « Vous êtes un analyste de données »</i>).
I	Instructions	Fournit des conseils étape par étape (<i>ex., « D'abord, résumez, puis comparez... »</i>).
C	Contexte	De quelles informations de base l'IA a-t-elle besoin ? (<i>par exemple, « Pour un public débutant » ou « Pour les applications de soins de santé »</i>).
A	Contraintes	Ajoute des règles supplémentaires (<i>par exemple, « Soyez concis », « Utilisez un langage formel », « Limitez la réponse à 150 mots »</i>).
	Additionnelles	

Lettre	Signification	Détail
R	Références	Des sources externes, des liens, des fichiers ou des données supplémentaires (<i>par exemple, « Utiliser cet ensemble de données » ou « Se référer à cet article »</i>).
D	Rendement Désiré	Spécifie le format de la réponse Gen AI (<i>par exemple, « créer un tableau avec 4 colonnes », « Fournir une liste à puces » ou « Sortie au format JSON »</i>).
O	Objectifs	Énonce clairement l'objectif final du prompt (<i>ex., « Rédigez un résumé de 100 mots » ou « Expliquez en termes simples »</i>).

Quand on utilise ce cadre, on est plus certain d'obtenir ce que l'on souhaite de l'IAG car on donne suffisamment d'informations sur le contexte et sur la structure de ce que l'on veut obtenir.

I.C. Le prompt à modifier

Pour faire votre analyse des besoins, vous avez une **base de prompt** où différents éléments ont déjà été écrits (*fichier texte fourni*). Les parties à compléter sont identifiées, vous pouvez rajouter quelques éléments mais pas modifier la structure du prompt ou enlever des parties.

Ce que vous devez rajouter correspond à votre choix de domaine d'activité, que vous devez préciser avec quelques informations. Identifiez si vous travaillez pour une entreprise, une organisation, une association, une administration... Rapprochez-vous de ce qui existe, trouvez des exemples desquels vous voulez vous rapprocher pour avoir un résultat précis. Par exemple, si vous voulez créer la BD d'une entreprise qui travaille dans la mode, vous pouvez vous inspirer de Louis Vuitton, Dior,...

Dans le prompt, vous devez ajouter des éléments qui correspondent au "Contexte" et aux "Références" (liens vers des sites web, des présentations, des articles,...) dans le framework RICARDO.

I.D. Analyse du résultat

L'IAG doit renvoyer le nombre de données attendues, et correspondant aux exigences que vous avez formulées dans votre prompt. En cas d'inadéquation avec votre prompt, apportez y des modifications et re-soumettez le **en entier**.

Si après plusieurs essais il y a toujours des erreurs dans votre résultat, vous êtes autorisé à enlever des données du dictionnaire s'il y en a trop. Indiquez alors les problèmes que vous avez relevés dans votre fichier de documentation **README.md**, et mentionnez les modifications que vous avez apportées.

Livrable

Le **prompt** final que vous avez utilisé, les **règles métier** et le **dictionnaire de données** obtenus doivent figurer dans votre fichier **README.md**. Le rapport et votre prompt au format texte sont à stocker dans votre répertoire Github.

Note : Vous êtes évalués sur votre capacité à écrire un prompt clair, précis et complet, et à vérifier l'adéquation du résultat à votre demande.

II. Deuxième étape : MCD

A partir du dictionnaire de données et des règles métiers obtenus à la première étape, produire le MCD qui doit respecter les contraintes suivantes :

1. Être produit avec un **outil de modélisation** (pas de schéma fait à la main).
2. Identifier clairement entités, associations, attributs, identifiants et cardinalités.
3. Respecter la **3FN**.
4. Il doit contenir des éléments de modélisation avancée vus en cours. C'est-à-dire **au moins 2 éléments** parmi :
 - une association récursive,
 - une association n-aire avec $n > 2$,
 - une entité faible et entité forte.Il faut que cette modélisation corresponde bien aux relations entre les données que vous voulez décrire, vous êtes évalués là-dessus.
5. Il doit intégrer **toutes les données** du dictionnaire obtenu, mais vous êtes libres d'en ajouter si vous le souhaitez.

Livrable

L'image de votre **MCD** doit figurer dans votre **README.md**, et le fichier source obtenu avec votre logiciel de modélisation est à ajouter à votre dépôt git.

Note : Vous êtes évalués sur l'adéquation du MCD au dictionnaire de données et aux règles métier, et l'utilisation appropriée de l'ensemble des éléments de modélisation.

Premier rendu

Pour le premier rendu, vous devez donc avoir **dans votre répertoire Github** :

- le prompt utilisé,
- ce que l'IA a renvoyé (règles métier et dictionnaire de données) dans un fichier de documentation README.md, qui doit être structuré pour décrire les différentes parties du projet,
- le fichier source du MCD.

Chaque membre du groupe doit avoir fait des commit vers le répertoire github.

Note : Ce travail servira de base pour appliquer les étapes suivantes et développer la base de données. Vous pourrez modifier votre MCD par la suite si vous identifiez des erreurs de modélisation après le rendu, mais c'est le MCD que vous aurez soumis lors du premier rendu qui sera évalué.

Références

- Sujet conçu et écrit par **Lena TREBAUL**, enseignante en informatique
- Framework pour le prompt engineering : **Salim NAHLE**, responsable des programmes Data & IA