

# Mini-project : Database design and development - PART 1

## Introduction

This mini-project covers all the steps of the MERISE method seen in this module and used in industry for database design and development. Based on a field that interests you, you will build your database by following these steps :

### Step 1 : Analyze requirements

- Choose a field that interests you
- Use the IAG to obtain information about the business rules of your field and the data to be stored

### Step 2 : MCD

- Using the data obtained in step 1, produce the MCD
- It must be normalized and include advanced modeling elements

### Step 3 : MLD and MLP

- From the MCD, retrieve the MLD and MPD in SQL
- Define additional constraints and the corresponding SQL code

### Step 4 : Insert data

- Populate the database using the IAG
- Ensure that the integrity constraints of the generated data are respected

### Step 5 : Query the database

- Imagine a scenario for using your database
- Write the SQL DML queries corresponding to this scenario

### Step 6 : Presentation

- Present your work in a video
- Explain your choices for modeling and using the database

- **Part 1** of the project corresponds to **steps 1 and 2**, and **Part 2** corresponds to **steps 3, 4, 5, and 6**.
- This project accounts for **20%** of the module grade.

- The project is to be carried out in **pairs**. Only one trio is allowed if there is an odd number of students.
- It is divided into two parts :
  - The topic for the first part of the project is made available on **February 13, 2026** and must be submitted by **February 27, 2025**.
  - The topic for the second part will be made available on **February 27, 2026** and must be submitted by **March 16, 2026**.
- You must submit (*non-exhaustive list, details are provided below*) :
  - a documentation file **README.md**,
  - various **text files** used for designing and inserting data (prompts),
  - **executable files** `.sql`,
  - a **video** presentation.
- All files must be stored in a **Github directory** shared with your lab instructor. There is nothing to send to them, *each submission is only a deposit in the Github directory*.

---

**Part 1** corresponds to the analysis of requirements and the creation of the MCD.

## I. First step : requirements analysis

### I.A. Gathering requirements

Requirements analysis involves obtaining all the necessary information about what needs to be stored. This information corresponds to a specific subject or domain. In the professional world, it is obtained by communicating with the company that wants to create its database.

In exercises, you have used descriptions that included business rules. Examples of business rules include : “An employee cannot be affiliated with more than one department,” “Employees have a manager,” and “A department has a name in the form ‘S’ + 2-digit number.”

For this project, your contact will not be a company/organization/administration, but an AI that will play the role of a company/organization/administration.

### I.B. Prompt engineering

To query the AI, you will use a prompt that uses an appropriate framework : the RICARDO framework, which allows you to write “good” prompts.

Letter	Meaning	Detail
R	Role	Defines the role of the generative AI (e.g., “Act as a teacher” or “You are a data analyst”).
I	Instructions	Provides step-by-step guidance (e.g., “First, summarize, then compare...”).
C	Context	What background information does the AI need ? (e.g., “For a beginner audience” or “For healthcare applications”).

Letter	Meaning	Detail
A	Additional Constraints	Adds additional rules (e.g., “Be concise,” “Use formal language,” “Limit the response to 150 words”).
R	References	External sources, links, files, or additional data (e.g., “Use this dataset” or “Refer to this article”)*.
D	Desired Output	Specifies the format of the Gen AI response (e.g., “create a table with 4 columns,” “provide a bulleted list,” or “output in JSON format”).
O	Objectives	Clearly states the ultimate goal of the prompt (e.g., “Write a 100-word summary” or “Explain in simple terms”).

When using this framework, you are more likely to get what you want from the LLM because you provide sufficient information about the context and structure of what you want to get.

### I.C. The prompt to be modified

To analyze your needs, you have a **prompt base** where various elements have already been written (*text file provided*). The parts to be completed are identified. You can add a few elements, but you cannot modify the structure of the prompt or remove any parts.

What you need to add corresponds to your choice of field of activity, which you must specify with some information. Identify whether you work for a company, an organization, an association, an administration, etc. Look at what already exists and find examples that you want to emulate in order to achieve a specific result. For example, if you want to create a comic book for a company that works in fashion, you can take inspiration from Louis Vuitton, Dior, etc.

In the prompt, you must add elements that correspond to the “Context” and “References” (links to websites, presentations, articles, etc.) in the RICARDO framework.

### I.D. Analysis of the result

The IAG must return the expected amount of data, corresponding to the requirements you specified in your prompt. If the result does not match your prompt, make changes and resubmit the **entire** prompt.

If, after several attempts, there are still errors in your result, you are allowed to remove data from the dictionary if there is too much. In this case, indicate the problems you have identified in your **README.md** documentation file and mention the changes you have made.

#### Deliverable

The final **prompt** you used, the **business rules**, and the **data dictionary** obtained must be included in your **README.md** file. The report and your prompt in text format should be stored in your Github directory.

**Note :** You will be assessed on your ability to write a clear, precise, and complete prompt, and to verify that the result matches your request.

## II. Second step : MCD

Using the data dictionary and business rules obtained in the first step, produce the MCD, which must comply with the following constraints :

1. Be produced with a **modeling tool** (no hand-drawn diagrams).
2. Clearly identify entities, relationship, attributes, identifiers, and cardinalities.
3. Comply with **3NF**.
4. It must contain advanced modeling elements seen in class. That is, **at least 2 elements** from among :
  - a recursive relationship,
  - an n-ary relationship with  $n > 2$ ,
  - a weak entity and a strong entity.

This modeling must correspond to the relationships between the data you want to describe ; you will be evaluated on this. 5. It must include **all the data** from the dictionary obtained, but you are free to add more if you wish.

### Deliverable

The image of your **MCD** must appear in your **README.md**, and the source file obtained with your modeling software must be added to your git repository.

**Note** : You will be assessed on the suitability of the MCD to the data dictionary and business rules, and the appropriate use of all modeling elements.

### First submission

For the first submission, you must therefore have **in your Github directory** :

- the prompt used,
- what the AI returned (business rules and data dictionary) in a README.md documentation file, which must be structured to describe the different parts of the project,
- the MCD source file.

Each member of the group must have made commits to the GitHub directory.

**Note** : This work will serve as the basis for applying the following steps and developing the database. You can modify your MCD later if you identify modeling errors after the submission, but it is the MCD that you submitted during the first submission that will be evaluated.

## References

- Topic designed and written by **Lena TREBAUL**, computer science teacher
- Framework for prompt engineering : **Salim NAHLE**, head of Data & AI Programs