ISAE SUPAERO
Institut Supérieur de l'Aéronautique et de l'Espace

TRIMANE
THE DATA INTELLIGENCE COMPANY

A I B T 102
DATA INTEGRATION & EXPLORATION

Study case

Bilel SDIRI, PhD

October 2022

Study case description

The aim of the study case is to put into practice and further study the different notions that have been presented as part of the course *AIBT102-Data Integration & Exploration*. More precisely, you are invited to plan and structure the different data workflow steps related to your study case. The application domain and context are to be set by the students. This should allow them to link the study with relevant business issues related to their traineeship companies, past professional experience, the targeted expertise level (functional/business, technical, or both), or focus on specific parts of the data integration and exploration flow.

The following points should be addressed while designing and implementing the study case:

a) Define the project that will be carried out and describe it conveniently. It is important to state clearly the **context**, the **business application domain**, the **business requirements/needs**, and the subsequent **relevant use cases**.

b) Identify clearly your data **sources** and **types**. This can be documented using a table that present for each data source, a description, the data type (i.e. structured, semi-structured, unstructured), and the link/website to download it or query data from it (i.e. in the case of an API for example). It is **recommended to diversity the sources and types** in order to explore and highlight the different challenges faced when collecting, processing, modeling, and analyzing your data.

c) Model your data: establish the **conceptual**, **logical**, and **physical** data models related to your study case. You can use *Vertabelo* or any other tool to draw your data models. In this step, it is important to **state clearly which modeling approach to adopt** (i.e. relational, dimensional) and **justify that**.

d) Illustrate (i.e. a schema) your data architecture, describe each step of your data architecture, and explain the motivations behind it. It is important also to **identify** and **describe** clearly the **data transformations** that will be performed in each step. With this in mind, it is useful to establish a **data dictionary** that lists the different attributes/column names, their source Tables/Collections, high-level description, rules/formulas/transformations.

e) Identify and describe **analytical** use-cases/needs/business-requirements/issues related to your applications. It is important, for each analytical case, to present a brief state-of-the-art regarding the possible solutions and justify the adopted methods/techniques. The analytical case can be related also to the way you organize, model, or transform your data as part of your data workflow. The solutions can be based on IA techniques, Data visualization, Programming, Data Modeling, or queries.

The deliverables of the study case are twofold:

1. A report detailing the points a-e.
2. The different technical materials developed/implemented as part of your study case (Scripts, programs, Colabs, Notebooks, ETL jobs, queries).