Technologies and Architecture for Data

- Hélène de Ribaupierre
- Spring Semester 2025
- S402017CR/SE
- M1150

Aim of the semester project

- Design and implement a data architecture solution for an analytical pipeline for a real-world dataset.
- The project should contain reflection about:
  - Data modelling
  - Storage
  - Integration/Transformation
  - Governance
  - Analysis
  - Evaluation (if needed)

The code:

- Working prototype to demonstrate key functionality
  - ETL
  - Analytics
  - Evaluation
  - Visualisation

Data Modelling (some questions) this is not exhaustive

- What are the data?
- What is the business case linked to the data? => model could change depending what we need to analyse
  - Reflect on what could be done differently depending on the use case
- What is the best model => conceptual one

Conceptual model is linked to the business case, but not to the database used

Storage (some questions) this is not exhaustive

- What is the appropriate storage?
- How do you store the data?
- Which format do you store the data?
- Which database type is the best for the use case (could be another type of storage)

Integration (some questions) this is not exhaustive

- How is the data ingested?
- What is the ETL/ELT?
- What are the transformation needed?
- What is the frequency of update?
- What is the quality of the data?

Governance (some questions) this is not exhaustive

- What are the issues you could face with your data?
- Security, Privacy?
- Who have access to your data?
- Can we delete the data?
- …

Design the data architecture

- High level architecture (possible schemas)
  - Data sources (what are their formats)
  - Storage solutions (how do you store the data)
  - Data flow and integration methods (ETL/ELT, APIs, streaming)
  - Governance and security considerations

Data model

- Design :
  - conceptual model
  - logical model (this will be more dependant of the technologies chosen)
  - using the appropriate notation

Technologies used

- What is the best technologies used for your business case and the data
  - Databases or data storage
  - ETL/ELT
  - Tools
  - Language
  - …

Tips

- Always compare and contrast with other possible choices (technologies, models, etc)
- Use scientific articles as references (when available)