

# Final Project – Data Infrastructure Automation & Deployment

---

Group Work: Teams of 5 students

Deadline:

- October 15th (11:59 PM) for **M2 #1**
- October 20th (11:59 PM) for **M2 #2**

## Context

As part of this professional simulation, you are asked to design and deploy a complete data infrastructure for a fictitious company. Your solution should take into account technical constraints, operational requirements, and compliance with current standards and best practices.

You may choose the type of company for which you are designing the infrastructure.

To inspire you, here are some example scenarios:

- E-commerce Platform: Handling user activity logs, recommendation engines, and large-scale transactions.
- Healthcare System: Managing electronic health records, AI-based diagnostics, and secure patient data.
- IoT Smart City Platform: Collecting and analyzing real-time sensor data from transportation, energy, and utilities.
- Financial Services (FinTech): Processing transactions, fraud detection with AI, and compliance monitoring.
- Streaming Platform: Managing user activity, video delivery performance, and recommendation systems.
- Logistics & Transportation: Optimizing routes, monitoring fleet data, and ensuring real-time analytics.

👉 You are free to create your own business case if you prefer.

## Objectives

1. Identify architecture requirements by analyzing constraints (technical, operational, regulatory).
2. Design a data architecture that fits business needs and ensures scalability, security, and compliance.
3. Deploy infrastructure components in the Cloud or On-Premise environment.
4. Orchestrate and monitor the infrastructure using industry-standard tools.

5. Document, demonstrate, and present the project to an evaluation jury.

## **Expected Deliverables**

### **1. Integration of Deliverables**

- Architecture requirements document (constraints, needs, design choices).
- Dockerized application + Kubernetes deployment.
- Cluster setup + benchmark results (performance, scalability).
- Airflow DAG orchestration (workflow automation).
- Grafana monitoring dashboard (system performance & health monitoring).

### **2. Documentation**

- Technical documentation in Markdown (GitHub Wiki/README).

### **3. Demonstration**

- Video demo: short recording of the infrastructure running in production (cloud or on-prem).

### **4. Jury Simulation**

- Oral defense in RNCP format:
- 10 min demo
- 15 min Q&A with jury

## **Additional Rules & Constraints**

- Architecture Design: Provide a complete architecture plan (PowerPoint, Google Slides, or equivalent diagram tool).
- Host deployment code on GitHub.
- Data Models: Include logical & physical data models.
- Monitoring & Optimization: Ensure proactive monitoring and troubleshooting with Grafana (or equivalent).
- Accessibility: Documentation and deliverables must remain clear and accessible (adapted formats if necessary).

## **Evaluation Criteria**

- Architecture relevance: Compliance with technical & regulatory requirements.
- Robustness & scalability: Ability to handle large-scale data and workloads.
- Deployment quality: Proper use of containerization, orchestration, and monitoring tools.
- Documentation clarity: Technical completeness and accessibility.
- Presentation & defense: Quality of oral presentation, demo, and answers during Q&A.

## Evaluation Process

- Presentation by the group: 10 min
- Q&A with jury: 15 min

## Final Deliverable

- Architecture plan
- Source code & configurations (on GitHub)
- Documentation in Markdown
- Video demo
- Oral defense presentation

Share it on Blackboard (or at [prillard.martin@gmail.com](mailto:prillard.martin@gmail.com))