

# Practical Work

## Big Data Laboratory

Informatics Engineering Master – Data Engineering  
2nd year 1st semester  
2024/2025

---

1. Objectives
  2. Schedule
  3. Regulations
    - 3.1 Scientific Paper
    - 3.2 Evaluation
  4. Work Description
  5. Bibliographic References
- 

### 1. Objectives

#### General Objective:

- Propose an architecture for a Big Data Project in different application domains.

#### Specific Objectives:

- Define the problem and describe a Big Data project to be analyzed
- Make a literature review on the problem and previous contributions to the domain of the selected Big Data project
- Select technologies and frameworks
- Propose an architecture prototype for the problem resolution (components, technologies and frameworks to be used on each component)
- Define the work methodology
- Writing a Technical Paper

## 2. Schedule

**Release of work proposals:** until October 4, 2024

**Work submission:** until **October 27, 2024** (23:59)

**Defense and discussion:** to be scheduled by the PL professor.

## 3. Rules

- The group (**maximum 3 members**) must be the same in both iterations of the Practical Work.
- Python should be used as the tool for data processing support.
- The deadline for **SUBMISSION** the Practical Work is **October 27, 2024**, in Moodle. Regardless of this deadline, groups should be able to report the development status of the Practical Work when requested by the teacher's.
- The Practical Work consists of a Scientific Paper.
- You should submit all documents in a compressed file. The zip file should contain:
  - the scientific paper in PDF format
  - datasets
  - presentation in ppt (10 minutes)
- The filename should follow the following notation:  
**LABGDD\_YYY\_XXX\_StudentNo1\_StudentNo2\_StudentNo3.zip**, where **YYY** represents the PL teacher's acronym, and **XXX** represents the PL class.

**Example: LABGDD\_AMD\_M1A\_7777777\_8888888\_9999999.zip.**

- Works whose designation does not respect the indicated notation **will be penalized by 10%.**
- **The practical work must be submitted on Moodle by the defined deadline. Late submissions will not be accepted.**
- The defense and discussion of the practical work will take place on a day and time to be scheduled by each Laboratory classes teacher. On the day of the presentation, **ALL** teamwork members must be present. Absent members will not receive a grade. The defense and discussion will be conducted as a team with questions directed to each individual member.
- Each teamwork is responsible for managing its development process. Difficulties and problems should be communicated to the laboratory class teacher in a timely manner.
- Code of conduct: (cf. IPP Student Disciplinary Regulations)
  - No student or teamwork can claim ownership of work done by others or developed in collaboration with third-parties.
  - The use of materials, artifacts, or code from others without proper and explicit indication of origin is expressly prohibited.
  - Code from other sources must be clearly identified within the code itself, indicating the source.
  - Cases of unauthorized appropriation of materials, artifacts, and/or code subject to assessment will be reported to the ISEP Dean.
  - The use of AI coding/drawing assistance tools (e.g. ChatGPT) must be mentioned.

- The use of the Bitbucket version control tool is mandatory.

### 3.1. Scientific Paper

The scientific paper should document all phases of the work methodology, problem definition, literature review, proposed architecture and conclusions (maximum of 8 pages using the IEEE template provided on Moodle). Consider the following aspects:

- The paper should have an initial section with a summary of the motivation and objectives of the work and a final section with a summary of the main conclusions drawn from the resolution of the different questions identified in the statement.
- Describe the components for the Big Data architecture development: problem definition and the Big Data project; literature review; selection of technologies and frameworks; Big Data architecture proposal. A synthesis of the developed work should be included in the final conclusions section.

### 3.2. Evaluation

In the evaluation of the Practical Work assessment, the following aspects and weights specified in Table 1 will be considered:

- a) Contextualization and objectives (Abstract and Introduction)
- b) Literature Review
- c) Significance and complexity of the dataset
- d) Organization, quality of writing, presentation, and clarity of the scientific paper
- e) Defense and discussion
- f) Individual participation of each group member

**Table 1 – Practical work Evaluation Grid**

Abstract and Introduction	15%
Problem definition	10%
Literature Review	25%
Big Data Architecture	35%
Conclusions and references	15%

**Note:** The grade of each group member will be determined based on their participation (in %). The practical work evaluation team will validate, at the time of the work defense (which may be through video conference), the participation of each group member in achieving the work and group objectives. **Absent members will not receive a grade.**

## 4. Work Description

With the development of the Practical Work, students are expected to develop skills in the development of approaches/solutions for Big Data projects [1-3], culminating in the writing of a scientific paper with the definition of the problem, a literature review, the proposed architecture and conclusions.

The primary objective is to foster skills in the development of Big Data architectures in different application domains.

## 5. Bibliographic References

- [1]. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016 (<https://www.deeplearningbook.org/>).

- [2]. P. Zikopoulos & C. Eaton, Understanding big data: Analytics for enterprise class hadoop and streaming data McGraw-Hill Osborne Media, 2011.
- [3]. B. Chambers & M. Zaharia, Spark: The definitive guide: Big data processing made simple, O'Reilly Media, Inc, 2018.- Online resources and documentation of Tensorflow, Hadoop, Pandas, Numpy, Apache Spark, and others.
- [4]. Online resources and documentation from Tensorflow (<https://www.tensorflow.org/overview>)