



**National University of Engineering (UNI)**  
School of Cybersecurity  
Syllabus 2024-II

**1. COURSE**

CS370. Big Data (Mandatory)

**2. GENERAL INFORMATION**

- 2.1 Course : CS370. Big Data
- 2.2 Semester : 9<sup>th</sup> Semester.
- 2.3 Credits : 3
- 2.4 Horas : 1 HT; 4 HP;
- 2.5 Duration of the period : 16 weeks
- 2.6 Type of course : Mandatory
- 2.7 Learning modality : Face to face
  - CS272. Databases II. (5<sup>th</sup> Sem)
- 2.8 Prerequisites :
  - CS3P1. Parallel and Distributed Computing . (8<sup>th</sup> Sem)

**3. PROFESSORS**

Meetings after coordination with the professor

**4. INTRODUCTION TO THE COURSE**

Nowadays, knowing scalable approaches to processing and storing large volumes of information (terabytes, petabytes and even exabytes) is fundamental in computer science courses. Every day, every hour, every minute generates a large amount of information which needs to be processed, stored, analyzed.

**5. GOALS**

- That the student is able to create parallel applications to process large volumes of information
- That the student is able to compare the alternatives for the processing of big data
- That the student is able to propose architectures for a scalable application

**6. COMPETENCES**

1) ()

6) Apply security principles and practices to maintain operations in the presence of risks and threats.()

**7. TOPICS**

Unit 1: Introducción a Big Data (15 hours)	
Competences Expected:	
Topics	Learning Outcomes
<ul style="list-style-type: none"><li>• Overview on Cloud Computing</li><li>• Distributed File System Overview</li><li>• Overview of the MapReduce programming model</li></ul>	<ul style="list-style-type: none"><li>• Explain the concept of Cloud Computing from the point of view of Big Data[Familiarizarse]</li><li>• Explain the concept of Distributed File System [Familiarizarse]</li><li>• Explain the concept of the MapReduce programming model[Familiarizarse]</li></ul>
Readings : [coulouris]	

Unit 2: Hadoop (15 hours)	
Competences Expected:	
Topics	Learning Outcomes
<ul style="list-style-type: none"> <li>• Hadoop overview.</li> <li>• History.</li> <li>• Hadoop Structure.</li> <li>• HDFS, Hadoop Distributed File System.</li> <li>• Programming Model MapReduce</li> </ul>	<ul style="list-style-type: none"> <li>• Understand and explain the Hadoop suite [Familiarizarse]</li> <li>• Implement solutions using the MapReduce programming model. [Usar]</li> <li>• Understand how data is saved in the HDFS. [Familiarizarse]</li> </ul>
Readings : [dongarra], [buyya]	

Unit 3: Procesamiento de Grafos en larga escala (10 hours)	
Competences Expected:	
Topics	Learning Outcomes
<ul style="list-style-type: none"> <li>• Pregel: A System for Large-scale Graph Processing.</li> <li>• Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud.</li> <li>• Apache Giraph is an iterative graph processing system built for high scalability.</li> </ul>	<ul style="list-style-type: none"> <li>• Understand and explain the architecture of the Pregel project. [Familiarizarse]</li> <li>• Understand the GraphLab project architecture. [Familiarizarse]</li> <li>• Understand the architecture of the Giraph project. [Familiarizarse]</li> <li>• Implement solutions using Pregel, GraphLab or Giraph. [Usar]</li> </ul>
Readings : [graphlab], [pregel], [giraph]	

## 8. WORKPLAN

### 8.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

### 8.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

### 8.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

## 9. EVALUATION SYSTEM

\*\*\*\*\* EVALUATION MISSING \*\*\*\*\*

## 10. BASIC BIBLIOGRAPHY