COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001

INTERNATIONAL
CAMPUS OF
EXCELLENCE

POLITÉCNICA

E.T.S. de Ingenieria y Sistemas
de Telecomunicacion

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

593000509 - Big Data Applications For Iot

DEGREE PROGRAMME

59AH - Master Universitario En Internet Of Things (iot)

ACADEMIC YEAR & SEMESTER

2022/23 - Semester 2

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	593000509 - Big Data Applications For Iot
No of credits	4.5 ECTS
Type	Compulsory
Academic year ot the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	59AH - Master Universitario en Internet Of Things (Iot)
Centre	59 - Escuela Tecnica Superior De Ingenieria Y Sistemas De Telecomunicacion
Academic year	2022-23

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Bonifacio Alberto Mozo Velasco (Subject coordinator)	ETSISI- 4313	a.mozo@upm.es	Sin horario.
Javier Martin Rueda		javier.martin@upm.es	Sin horario.
Maria Dolon Poza	ETSIST 8303	maria.dolonp@upm.es	Sin horario.

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

The subject - recommended (passed), are not defined.

3.2. Other recommended learning outcomes

- There is no entry requirements although it is highly recommended to have solid foundations on programming languages (such as C, Java or Python) and data structures

4. Skills and learning outcomes *

4.1. Skills to be learned

CB07 - Students should be able to apply their acquired knowledge and problem-solving skills in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study.

CB09 - Students are able to communicate their conclusions and the knowledge and rationale behind them to specialist and non-specialist audiences in a clear and unambiguous way.

CE.11 - Design applications capable of processing large volumes of information and data streams used in IoT applications including the storage, management and analysis of available information.

CG02 - Students will be able to apply advanced methods and technologies to address needs and problems in IoT applications.

CG04 - Students will have the ability to apply efficiency, scalability, reliability and security criteria in different areas of smart applications and cyber-physical systems, such as Smart Living, Smart Cities or eHealth.

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

CT.01 - Ability to use English for work in international contexts.

CT.03 - Creativity, initiative and entrepreneurship

CT.04 - Ability to design, plan, coordinate and manage projects technically and economically according to ethical, quality and environmental criteria.

4.2. Learning outcomes

RA36 - To extract information and knowledge for an organization from a massive data set

RA34 - To apply the most appropriate management and analysis techniques for a massive data set

RA44 - To select the most appropriate architecture and storage for a massive set of heterogeneous data within the scope of IoT

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

This course introduces the student to the massive data processing problem. Different paradigms will be described considering both batch and online strategies.

Some of the most popular distributed computing and data processing platforms in the current Big Data scenario will be presented and studied: Apache Hadoop, Apache Spark, and the Elastick ELK stack.

Additionally, the student will learn the foundations of Machine Learning techniques and how to apply them in Big Data Use Cases.

5.2. Syllabus

1. Introduction to Big-Data: concepts, challenges and models

2. The Elastic Stack

3. Introduction to Distributed Computing: MapReduce and Hadoop

4. Distributed Computng with Apache Spark. RDDs and the Spark API

5. Introduction to Machine Learning

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	**Introduction to Big-Data** Duration: 03:30			
2	**The Elastic Stack** Duration: 02:30			**Test #1 Big Data and Elastic Stack Concepts** Continuous assessment and final examination Presential Duration: 01:00
3	**The Elastic Stack** Duration: 02:30			**Test #2 Activity 1 and Logstash** Continuous assessment and final examination Presential Duration: 01:00
4	**Introduction to Distriibuted Computing: Map Reduce and Hadoop** Duration: 01:00	**Parallelization of applications using Map-Reduce** Duration: 02:30		
5	**Introduction to Distriibuted Computing: Map Reduce and Hadoop** Duration: 01:00	**Big Data applications using Hadoop** Duration: 02:30		
6		**Big Data applications using Hadoop** Duration: 02:30		**Test #3 Hadoop** Continuous assessment and final examination Presential Duration: 01:00
7	**Distributed Computing with Apache Spark. RDDs and the Spark API** Duration: 01:00	**Parallelization of applications using Apache Spark** Duration: 02:30		
8		**Parallelization of applications using Apache Spark** Duration: 03:30		**Parallelization of Big Data applications with Apache Spark** Continuous assessment and final examination Presential Duration: 00:00
9	**Introduction to Machine Learning** Duration: 03:30			

10		**Parallel implementation in Spark of a ML classifier** Duration: 03:30		
11		**Parallel implementation in Spark of a ML classifier** Duration: 03:30		
12		**Parallel implementation in Spark of a ML classifier** Duration: 01:30		**Parallel implementation in Spark of a Spam Classifier** Continuous assessment and final examination Presential Duration: 00:00 **Part #2. Individual Exam Units 4 and 5** Continuous assessment and final examination Presential Duration: 02:00
13				
14				
15				
16				
17				

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
2	Test #1 Big Data and Elastic Stack Concepts		Face-to-face	01:00	16.65%	5 / 10	CB07 CT.01 CG02 CE.11
3	Test #2 Activity 1 and Logstash		Face-to-face	01:00	16.65%	5 / 10	CB07 CT.01 CG02 CE.11
6	Test #3 Hadoop		Face-to-face	01:00	16.7%	5 / 10	CB07 CT.01 CG02 CE.11
8	Parallelization of Big Data applications with Apache Spark		Face-to-face	00:00	10%	5 / 10	CB07 CT.01 CG02 CT.03 CE.11
12	Parallel implementation in Spark of a Spam Classifier		Face-to-face	00:00	25%	5 / 10	CB07 CG02 CT.03 CE.11
12	Part #2. Individual Exam Units 4 and 5		Face-to-face	02:00	15%	5 / 10	CB07 CT.01 CT.04 CG04 CB09 CT.03 CE.11

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills

2	Test #1 Big Data and Elastic Stack Concepts		Face-to-face	01:00	16.65%	5 / 10	CB07 CT.01 CG02 CE.11
3	Test #2 Activity 1 and Logstash		Face-to-face	01:00	16.65%	5 / 10	CB07 CT.01 CG02 CE.11
6	Test #3 Hadoop		Face-to-face	01:00	16.7%	5 / 10	CB07 CT.01 CG02 CE.11
8	Parallelization of Big Data applications with Apache Spark		Face-to-face	00:00	10%	5 / 10	CB07 CT.01 CG02 CT.03 CE.11
12	Parallel implementation in Spark of a Spam Classifier		Face-to-face	00:00	25%	5 / 10	CB07 CG02 CT.03 CE.11
12	Part #2. Individual Exam Units 4 and 5		Face-to-face	02:00	15%	5 / 10	CB07 CT.01 CT.04 CG04 CB09 CT.03 CE.11

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
The student will take all individual evaluations of Units 1,2 and 3		Face-to-face	01:30	50%	5 / 10	CB07 CT.01 CT.04 CG04 CB09 CG02 CT.03 CE.11

The student wil hand the practical assignments of part 2 up to the day of the written exam		Face-to-face	00:00	35%	5 / 10	CB07 CT.01 CG04 CB09 CG02 CT.03 CE.11
The student will take an individual written exan covering Units 4 and 5		Face-to-face	02:00	15%	5 / 10	CB07 CT.01 CT.04 CG04 CB09 CG02 CT.03 CE.11

7.2. Assessment criteria

Continuous Assessment, Final examination and Referred (re-sit)examination

Part1

----

3 Individual evaluations

Final grade Part 1 = (sum of marks) / Number of evaluations

Part2

----

2 practical assignments and 1 individual written exam

N1= Individual Exam

P21= Practical assignment 1

P22= Practical assignment 2

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

Final grade Part 2 = 0.3*N1 + 0.2*P21 + 0.5*P22

Final grade= 0.5* Part1 + 0.5 * Part2

Assesment criteria: As a general criterion to pass the course, it is required that each evaluation is greater o equal than 5.0

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Course slides	Bibliography	Course slides

PR/CL/001
COORDINATION PROCESS OF
LEARNING ACTIVITIES

ANX-PR/CL/001-01
LEARNING GUIDE

E.T.S. de Ingenieria y
Sistemas de
Telecomunicacion

9. Other information

9.1. Other information about the subject

Sustainable Development Goals:

This subject is aligned with Goal 4 Quality Education that aims to "Guarantee inclusive, equitable and quality education and promote lifelong learning opportunities for all" that defined within the Sustainable Development Goals (SDGs) established by the United Nations Organization (UN).

Other considerations:

In anticipation of possible recurrences of the COVID epidemic and in the event of a possible suspension of all face-to-face educational activities, the exceptional procedure that will be carried out will be that the evaluation activities initially planned as face-to-face activities could be developed through online platforms. The online platforms to be used will be those provided by the University, both those accessible through Moodle and those available as a telematic resource.