



INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingenieros
Industriales

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

53001521 - Tools For Big Data Analytics

DEGREE PROGRAMME

05BD - Master Universitario En Ingenieria De La Organizacion

ACADEMIC YEAR & SEMESTER

2021/22 - Semester 1

Index

Learning guide

1. Description.....	1
2. Faculty.....	1
3. Prior knowledge recommended to take the subject.....	2
4. Skills and learning outcomes	2
5. Brief description of the subject and syllabus.....	4
6. Schedule.....	7
7. Activities and assessment criteria.....	9
8. Teaching resources.....	11
9. Other information.....	12

1. Description

1.1. Subject details

Name of the subject	53001521 - Tools For Big Data Analytics
No of credits	3 ECTS
Type	Optional
Academic year of the programme	Second year
Semester of tuition	Semester 3
Tuition period	September-January
Tuition languages	English
Degree programme	05BD - Master Universitario en Ingenieria de la Organizacion
Centre	05 - Escuela Tecnica Superior De Ingenieros Industriales
Academic year	2021-22

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Joaquin Bienvenido Ordieres Mere (Subject coordinator)	UD Proyectos	j.ordieres@upm.es	M - 08:30 - 14:30

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

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

- Bigdata: De Los Datos A Las Decisiones
- Tecnología, Organización Y Nuevos Modelos De Negocio Digital
- Métodos Cuantitativos Avanzados

3.2. Other recommended learning outcomes

- Programming Python
- Knowledge about Operating Systems such as Unix or Linux at user level

4. Skills and learning outcomes *

4.1. Skills to be learned

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

CE01 - Conocer y aplicar técnicas y herramientas para el manejo y análisis de grandes masas de datos

CG02 - Analizar situaciones estructuradas y poco estructuradas de empresas y otras organizaciones, estableciendo diagnósticos apropiados, en particular, de carácter estratégico

CG03 - Concebir soluciones para afrontar problemas previamente diagnosticados, y evaluarlas desde diferentes criterios correspondientes a los distintos actores concernidos

CT01 - Aplica. Habilidad para aplicar conocimientos científicos, matemáticos y tecnológicos en sistemas relacionados con la práctica de la ingeniería

CT02 - Experimenta. Habilidad para diseñar y realizar experimentos así como analizar e interpretar datos

CT03 - Diseña. Habilidad para diseñar un sistema, componente o proceso que alcance los requisitos deseados teniendo en cuenta restricciones realistas tales como las económicas, medioambientales, sociales, políticas, éticas, de salud y seguridad, de fabricación y de sostenibilidad

CT05 - Resuelve. Habilidad para identificar, formular y resolver problemas de ingeniería

CT08 - Entiende los impactos. Educación amplia necesaria para entender el impacto de las soluciones ingenieriles en un contexto social global

CT09 - Se actualiza. Reconocimiento de la necesidad y la habilidad para comprometerse al aprendizaje continuo

CT10 - Conoce. Conocimiento de los temas contemporáneos

CT11 - Usa herramientas. Habilidad para usar las técnicas, destrezas y herramientas ingenieriles modernas necesarias para la práctica de la ingeniería

CT12 - Es bilingüe. Capacidad de trabajar en un entorno bilingüe (inglés/español)

4.2. Learning outcomes

RA12 - Comprender y evaluar los efectos de la incorporación de las Tecnologías de la Información y las Comunicaciones (TIC) en todos los niveles de la organización

RA1 - Elegir y aplicar técnicas de análisis exploratorios de grandes bases de datos

RA13 - Identificar y analizar el impacto que las tecnologías digitales, de la comunicación y los datos y de la inteligencia artificial tienen en la aparición de nuevos modelos de negocio disruptivos

* 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 is taught in the third semester of the master in Organizational Engineering, in the fall season. As a general idea it aims to,

- Understand the real meaning of Big Data problems and solutions, through different use cases, to compare with classical IT solutions regularly found
- Be aware and able to apply ETL solutions for those environments.
- Be aware of the CI/CD contexts and microservice oriented platforms.
- Be able to create models and visualize the forecasts connecting them with decision making processes

The course is mainly oriented by case enrollment.

In previous editions this course took advantage of RosettaHub middleware to interface with AWS, making it possible to have a context for distributing computing. However, the UPM authorities did not continue the agreement with RosettaHub, therefore this solution is no longer available. Therefore, students are kindly requested to install anaconda on their own computers and practical knowledge regarding Hadoop/Spark will be provided by the instructor but no hardware will be granted for students to play with.

Application cases will be gathered from the real life, including datasets and other referential characteristics. The motivation for the analysis to be carry out will emphasize their contribution to the SDGs, highlighting how these technologies can help organizations to do better regarding such aspects.

Deeply, the SDGs to be worked with, are:

SDG8: Good Jobs and Economic Growth. Because of students will realize how value creation can be linked to the

successful contribution to solve the raised cases due to their effects

SDG9: Industry, Innovation and Infrastructure. Most of the examples will be linked to the industry examples, as cases allow to observe effects from the shopfloor to the organizational level as well.

SDG3: Good Health and Well being. Because of workers are key aspects in the value stream of the organizations and cases will care about sensing them.

5.2. Syllabus

1. INTRODUCTION AND BUSINESS PERSPECTIVE

- 1.1. Why do we need big data solutions? Who are using such approaches?
- 1.2. What does IoT and IIoT mean ? What is the foreseen impact in regular business?
- 1.3. How to store data in BD? How does it work? A Technology perspective

2. BIG-DATA ARCHITECTURE

- 2.1. Clusters
- 2.2. Data bases in BigData

3. BIGDATA ANALYTICS

- 3.1. Doing operations with data living in BD environments
- 3.2. Spark solution and other related stuff around

4. COMMERTIAL vs HOME MADE SOLUTIONS

- 4.1. The big players
- 4.2. OpenSource solutions

5. DEEP LEARNING

- 5.1. What does DL stand for?
- 5.2. Is it the same that Machine Learning?
- 5.3. Creating ML models in python

5.4. Creating DL models in python with Keras and TF

6. PRACTICAL ASSIGNMENT

6.1. CRISP/DM Framework

6.2. Dealing with the data and the business case

6.3. Growing models

6.4. Analysis of results

6.5. Value proposal

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 the course. Lecture 1 Duration: 02:00 Lecture			
2	Lecture 1 Duration: 02:00 Lecture			Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05
3	Architecture(s) analysis and discussion Duration: 01:00 Lecture	Case selection to be developed during the course. Duration: 01:00 Laboratory assignments		Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05
4	Big Analytics analysis and discussion Duration: 01:00 Lecture	Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments		Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05
5	Implementation strategies: Isolated, Distributed, CD/CI Duration: 01:00 Lecture	Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments		Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05
6	Refinement of the Group/Individual assignment Duration: 01:00 Cooperative activities	Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments		Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05
7		Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments	Effective development of the individual/group assignments Duration: 01:00 Additional activities	Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05
8		Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments	Effective development of the individual/group assignments Duration: 01:00 Additional activities	Gamification activity to assess the knowledge gathered. Online test Continuous assessment Presential Duration: 00:05

9		Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments	Effective development of the individual/group assignments Duration: 01:00 Additional activities	
10		Case development. Practical activity following the CRISP/DM methodology. Duration: 01:00 Laboratory assignments	Effective development of the individual/group assignments Duration: 01:00 Additional activities	
11		Case development. Practical activity following the CRISP/DM methodology. Duration: 02:00 Laboratory assignments		
12		Case development. Practical activity following the CRISP/DM methodology. Duration: 02:00 Laboratory assignments		
13		Case development. Practical activity following the CRISP/DM methodology. Duration: 02:00 Laboratory assignments		
14		Case development. Practical activity following the CRISP/DM methodology. Duration: 02:00 Laboratory assignments		Individual delivery of the practical assignment. Individual work Final examination Not Presential Duration: 00:30
15				Presentation of the achievements Group presentation Continuous assessment Presential Duration: 02:00 Delivery of the video presentig the strenghts and weaknesses of the work carried out. Individual presentation Final examination Not Presential Duration: 00:30
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.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
2	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
3	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
4	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
5	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
6	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
7	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
8	Gamification activity to assess the knowledge gathered.	Online test	Face-to-face	00:05	1%	3 / 10	CB06 CT12 CE01
15	Presentation of the achievements	Group presentation	Face-to-face	02:00	93%	5 / 10	CB06 CT03 CE01

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
14	Individual delivery of the practical assignment.	Individual work	No Presential	00:30	75%	5 / 10	CB06 CT03 CE01

15	Delivery of the video presentig the strenghts and weaknesses of the work carried out.	Individual presentation	No Presential	00:30	25%	5 / 10	CB06 CT12 CT03
----	---	-------------------------	---------------	-------	-----	--------	----------------------

7.1.3. Referred (re-sit) examination

No se ha definido la evaluación extraordinaria.

7.2. Assessment criteria

Group Assignments

The development of a group oriented practical assignment is strongly recommended, as well as its public presentation. Main reasons are connected to the SDG process of improvement as well as the required effort for addressing the different engagements are larger than availabilities from individuals. Score from this activity will be taken into account.

The topic for the group assignemnt can be suggested either for the instructor or for the students, However, in the last case the instructor shall agree upon the proposal.

Diversity in topics will be promoted, as well as data availability, either from public sources or provided by some stakeholder.

Under specific circumstances the instructor can share scraper, when convenient.

The assessment will always follow 360° and development will be under the Agile methodology, as much as possible.

Scoring Rules

There are two main alterantive approaches, where the student will select the most convenient one,por el alumno:

Continuous Assessment (Bologna style):

- Individual test for knowledge (based on the topics worked out during the course).
- Score for the group assignment, where each team member will score contributions from other group members.

The final score is composed by:

- Score from the individual test, while it is over 3 points: Weight = 10%
- Score from the group Assignment: Weight 80% (60% from instructor + 20% from Students)
- Additional bonus from the instructor (when applicable because of commitment, attendance, etc.): 10%

In case of failure to pass the Continuous Assessment method, students are kindly requested to attend the Final test.

Final Test (Tradicional style):

- Score from the individual test, while it is over 4 points: Weight = 25%
- Individual Practical Assignment (50%)
- Self explanatory video with the critical assessment of the findings (25%)

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Data-Intensive Text Processing with MapReduce by Jimmy Lin and Chris Dyer Morgan & Claypool Publishers, 2010.	Bibliography	
Hadoop Real World Solutions Cookbook by Jonathan R. Owens, Brian Femiano, and Jon Lentz Publication Date: February 7, 2013 ISBN-10: 1849519129 ISBN-13: 978-1849519120	Bibliography	

BlueMix	Equipment	Plataforma de uso de IBM
Amazon	Equipment	Plataforma de desarrollo
GCP	Equipment	Plataforma de desarrollo
Aula	Equipment	Se precisa aula informática y / o equipamiento informático para estas clases
python	Web resource	
keras	Web resource	
jupyter	Web resource	

9. Other information

9.1. Other information about the subject

As the course is handled in a practical way, the usage of computing resources is enforced. Therefore, it is expected the students bring their own computers.

The course is related to the following SDGs:

SDG8: Good Jobs and Economic Growth. Because of students will realize how value creation can be linked to the successful contribution to solve the raised cases due to their effects

SDG9: Industry, Innovation and Infrastructure. Most of the examples will be linked to the industry examples, as cases allow to observe effects from the shopfloor to the organizational level as well.

SDG3: Good Health and Well being. Because of workers are key aspects in the value stream of the organizations and cases will care about sensing them.