

TECHNICAL SHEET OF THE SUBJECT

Data of the subject	
Subject name	Big Data Technology
Subject code	DTC-BA-517
Main program	Grado en Análisis de Negocios/Business Analytics
Involved programs	Grado en Ingeniería en Tecnologías de Telecom. y Grado en Análisis de Negocios/Business Analytics [Fifth year]
Level	Reglada Grado Europeo
Quarter	Semestral
Credits	6,0 ECTS
Type	Obligatoria (Grado)
Department	Department of Telematics and Computer Sciences
Coordinator	Jorge López-Malla Matute

Teacher Information	
Teacher	
Name	Jorge López-Malla Matute
Department	Department of Telematics and Computer Sciences
E-Mail	jlopezmalla@icai.comillas.edu

SPECIFIC DATA OF THE SUBJECT

Contextualization of the subject
<p>Contribution to the professional profile of the degree</p> <p>During the course, the student will learn about distributed computing, the Hadoop ecosystem, NOSQL databases and Big Data technologies, both Batch and Streaming, focusing on Apache Spark and Apache Kafka.</p> <p>Once these concepts have been acquired, the student should be able to identify the needs of a distributed architecture depending on different uses cases also Will be able to solve different Big Data uses cases using both Map & Reduce and Apache Spark in no streaming uses cases and Apache Spark and Apache Kafka for the streaming ones.</p> <p>The student also must also be able to identify the best NOSQL storage solution for each use case.</p>
<p>Prerequisites</p> <p>Big Data</p>

Competencies - Objectives
<p>Competences</p>



GENERALES		
CG01	Capacidad de organización y planificación en la identificación de problemas en el contexto de datos masivos	
	RA1	Identificar y organizar temporalmente las tareas necesarias para la realización sus actividades de aprendizaje, cumpliendo con los plazos establecidos de entrega de actividades.
	RA2	Ser capaz de integrarse y participar en el desarrollo organizado de un trabajo en grupo.
CG02	Capacidad de análisis de datos masivos procedentes de diversas fuentes: texto, audio, numérica e imagen	
	RA1	Describir, relacionar e interpretar situaciones y planteamientos sencillos
	RA2	Seleccionar las fuentes de información digital más significativas
	RA3	Identificar las fuentes de información más relevantes para cada problema
	RA4	Utilizar las herramientas digitales idóneas para la adquisición y gestión de la información y es capaz de integrar esa información de una manera racional
CG03	Resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos	
	RA1	Resolver problemas y realizar trabajos aplicando una diversidad metodológica de análisis
	RA2	Identificar correctamente los conocimientos aplicables a cada situación
	RA3	Identificar la metodología más apropiada para la resolución de los problemas planteados, conociendo las herramientas software más relevantes y aportando una visión crítica.
CG07	Capacidad de liderazgo y trabajo en equipo, en la sociedad de la información	
	RA1	Participar de forma activa en el trabajo de grupo compartiendo información, conocimientos y experiencias
	RA2	Orientarse a la consecución de acuerdos y objetivos comunes
	RA3	Ser capaz de organizar y/o coordinar equipos de trabajo
CG08	Capacidad crítica y autocrítica en la sociedad de la información	
	RA1	Identificar los supuestos y las limitaciones de los modelos y metodologías a aplicar en cada situación
	RA2	Ejercer una valoración crítica de los modelos y metodologías a aplicar en cada situación
CG09	Compromiso ético en la sociedad de la información	
	RA1	Ser honesto en el desarrollo de las actividades académicas y en otros aspectos de la vida y no ser pasivo ante la deshonestidad de otros.
	RA2	Conocer y desarrollar las bases de la ética profesional



CG11	Capacidad para aprender y trabajar autónomamente en la sociedad de la información	
	RA1	Organizar los materiales de la asignatura con el fin de articular su aprendizaje
	RA2	Poseer criterio para decidir la utilización de los medios humanos y otros recursos puestos a su disposición para el aprendizaje
	RA3	Realizar sus trabajos y su actividad necesitando sólo unas indicaciones iniciales y un seguimiento básico
ESPECÍFICAS		
CE23	Conocer y utilizar las bases de datos relacionales	
	RA1	Conocer el concepto de base de datos y su importancia para el almacenamiento ordenado de la información
	RA2	Comprender y aplicar el concepto de entidad relación y es capaz de diseñar bases de datos relacionales
	RA3	Conocer la sintaxis básica de los sistemas de recuperación de información más utilizados
CE24	Conocer y aplicar técnicas más efectivas de visualización de datos y su importancia para la extracción de conocimiento	
	RA1	Conocer los tipos de representaciones más importantes y su idoneidad en función del tipo de información.
	RA2	Identificar los parámetros más relevantes en un conjunto de datos mediante técnicas elementales de análisis exploratorio
	RA3	Ser capaz de producir representaciones gráficas que sintetizan conjuntos de datos voluminosos
CE25	Comprender los conceptos más importantes asociados al Big Data y sus elementos característicos	
	RA1	Conocer el concepto de paralelismo y la importancia de la distribución de datos voluminosos
	RA2	Identificar las necesidades específicas asociadas al manejo, adquisición y almacenamiento de volúmenes masivos de datos.
	RA3	Comprender y conocer las tecnologías de comunicaciones (redes) y computación precisas para la distribución y recuperación de información, así como de su análisis.
CEO33	Conocer y usar técnicas de análisis de información para extraer el conocimiento para la toma de decisiones de negocio.	
	RA01	Conoce y comprende las herramientas de apoyo a la decisión en el ámbito de la empresa
	RA02	Es capaz de implementar una solución computerizada de apoyo a la decisión.
	RA03	Comprende y aplica herramientas de adquisición, representación e inferencia, y aquellas técnicas de aprendizaje automático y basadas en sistemas inteligentes



CEO34	Entender la computación y las bases de datos distribuidas, con énfasis en la arquitectura Hadoop y las herramientas actuales más extendidas en el contexto de Big Data.	
	RA01	Conoce y comprende la problemática del Big Data y las técnicas más avanzadas de computación y almacenamiento usadas en la actualidad.
	RA02	Comprende y es capaz de administrar una arquitectura Hadoop y otras arquitecturas NoSQL.
	RA03	Comprende y aplica las técnicas de MapReduce y otras técnicas avanzadas de computación específicas para el análisis de datos voluminosos.
CEO35	Comprender y ser capaz de desarrollar herramientas de visualización avanzadas.	
	RA01	Conoce y aplica las metodologías y herramientas de desarrollo de aplicaciones orientadas a la visualización.
	RA02	Identifica las necesidades de síntesis de información y la representación visual automatizada.
CEO36	Conocer y comprender las herramientas más extendidas de gestión de datos.	
	RA01	Conoce y comprende el modelo de extracción, transformación y carga de datos orientados a la gestión de la empresa.
	RA02	Conoce las tecnologías OLAP y su uso en el contexto de los datawarehouses corporativos.
	RA03	Conoce y es capaz de generar informes basados en la gestión de datos para la toma de decisiones.
CEO37	Conocer y aplicar las técnicas de extracción de información de redes sociales y portales web.	
	RA01	Conoce y comprende las características de las distintas redes sociales y la importancia de su análisis para la extracción de información relevante para la empresa
	RA02	Conoce y es capaz de administrar un sistema de recopilación, agregación y evaluación del <i>¿sentimiento¿</i> asociado a la imagen corporativa en redes sociales
	RA03	Es capaz de analizar métricas para el análisis del tráfico web de un portal así como la cuantificación de la experiencia de usuario de un servicio web.

THEMATIC BLOCKS AND CONTENTS

Contents - Thematic Blocks

1. Introduction to Big Data
2. Hadoop ecosystem
3. Distributed Storage technologies
4. Apache Spark: Uses cases and architecture
5. Machine Learning and Apache Spark

6. Streaming use cases in the Big data ecosystem
7. Spark Streaming
8. Apache Kafka
9. Datawarehouse and Data lake
10. Cloud technologies for Big Data use cases

TEACHING METHODOLOGY

General methodological aspects of the subject

SUMMARY STUDENT WORKING HOURS

CLASSROOM HOURS
NON-PRESENTIAL HOURS
ECTS CREDITS: 6,0 (0 hours)

EVALUATION AND CRITERIA

Evaluation activities	Evaluation criteria	Weight
Final exam	The final exam will have both test and practical exercises related to the different parts of the course	60 %
Hands-on exercises	Hands-on exercises using the different technologies explained during the course	10 %
Hands-on exercises	Hands-on exercises based on the different technologies explained during the course	10 %
Hands-on exercises	Hands-on exercises based on the different technologies explained during the course	10 %
Middle term exam	Middle term exam with both practical and test exercises	10 %

Ratings

The final grade (FG) will be calculated using the following formula:

$$NF = 50\% \text{ Final exam} + 10\% \text{ middle term exam} + 30\% \text{ Hand-on exercises} + 10\% \text{ attitude}$$

The minimum grades for each part are:

- Middle term: 4.
- Final exam: 5.
- Hands-on: 5

BIBLIOGRAPHY AND RESOURCES

In compliance with current regulations on the **protection of personal data**, we would like to inform you that you may consult the aspects related to privacy and data that you have accepted on your registration form by entering this website and clicking on "download"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>