

CENTRO UNIVERSITARIO DE TECNOLOGÍA Y ARTE DIGITAL



PLANIFICACIÓN DE LA DOCENCIA UNIVERSITARIA

GUÍA DOCENTE

PROCESAMIENTO DE DATOS

1. DATOS DE IDENTIFICACIÓN DE LA ASIGNATURA.

Título:	Grado en Ingeniería del Software
Facultad:	Centro Universitario de Tecnología y Arte Digital (U-TAD)
Materia:	Ingeniería de Datos
Denominación de la asignatura:	Procesamiento de Datos
Curso:	3º
Cuatrimestre:	2
Carácter:	Obligatoria de mención
Créditos ECTS:	6
Modalidad/es de enseñanza:	Híbrido Presencial
Idioma:	Castellano
Profesor/a:	Manoel Fernando Alonso Gadi
E-mail:	Manoel.alonso@u-tad.com
Teléfono:	

2. DESCRIPCIÓN DE LA ASIGNATURA.

2.1 Descripción de la materia

Los contenidos de la materia permiten a los alumnos comprender el flujo de búsqueda, ingesta, almacenamiento, procesamiento y análisis de información de datos y aproxima a los alumnos a las técnicas y tecnologías necesarias para la gestión de grandes cantidades de datos.

2.2 Descripción de la asignatura

El objetivo de este curso es entrar en el mundo de Big Data en un entorno distribuido y en real time.

Aprender procesamiento de datos distribuido es algo fundamental hoy en día, en este curso utilizaremos Spark. Spark es la tecnología que está revolucionando el mundo de la analítica y el big data. Spark es un motor de procesamiento de datos de código abierto creado en torno a la velocidad, la facilidad de uso y el análisis.

Empezaremos el curso con una Introducción a Python con el objetivo de utilizar Spark usando PySpark. Veremos conceptos básicos de Spark como tareas distribuidas, RDD y al arquitectura Máster/Slave y a continuación exploraremos los módulos/extensiones Spark SQL, Spark Streaming y Machine Learning con MLib en Spark.

3. COMPETENCIAS

Competencias Básicas y Generales

CB1: Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.

CB2: Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.

CB3: Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética.

CB4: Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.

CB5: Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CG1 - Capacidad para entender, planificar y resolver problemas a través del desarrollo de soluciones informáticas.

CG3 - Conocimiento de los fundamentos científicos aplicables a la resolución de problemas informáticos

CG4 - Capacidad para simplificar y optimizar los sistemas informáticos atendiendo a la comprensión de su complejidad

CG9 - Capacidad para aprender, modificar y producir nuevas tecnologías informáticas

CG10 - Uso de técnicas creativas para la realización de proyectos informáticos

CG11 - Capacidad de buscar, analizar y gestionar la información para poder extraer conocimiento de la misma

Competencias Específicas

CE3 - Conocimiento del álgebra relacional y realización de consultas en lenguajes procedurales para el diseño de esquemas de

bases de datos normalizados basados en modelos de entidad-relación

CE10 - Capacidad para manejar un gestor de versiones de código y generar la documentación de una aplicación de forma automática.

Competencias Transversales

CT1 - Conocimiento de la definición, el alcance y la puesta en práctica de los fundamentos de las metodologías de gestión de proyectos de desarrollo tecnológico

CT2 - Conocimiento de los principales agentes del sector y del ciclo de vida completo de un proyecto de desarrollo y comercialización de contenidos digitales

CT4 - Capacidad de actualización del conocimiento adquirido en el manejo de herramientas y tecnologías digitales en función del estado actual del sector y de las tecnologías empleadas

CT5 - Desarrollo de las habilidades necesarias para el emprendimiento digital.

4. CONTENIDOS

Tema 1.- Introducción a Spark y Python

Introducción y instalación de Spark. Introducción a las estructuras de datos en Python: listas, diccionarios y data frames. Métodos de manipulación de estructuras: lista comprendida, funciones anónimas/lambda y vía map/reduce/filter, apply y np.vectorize.

Tema 2.- Spark Básico

Conceptos básicos de Spark. Spark Core, tareas distribuidas, programación y funcionalidades básicas de I/O y RDD (Resilient Distributed Datasets).

Tema 3.- Spark Cluster

Arquitectura Máster/Slave - esclavos/trabajadores en el caso de Spark. El controlador y los ejecutores ejecutan sus procesos individuales y los usuarios pueden ejecutarlos en el mismo clúster de Spark o en máquinas separadas.

Tema 4.- Spark SQL

Spark SQL es un módulo Spark para el procesamiento de datos estructurados. Proporciona una abstracción de programación llamada DataFrames y también puede actuar como un motor de consultas SQL distribuido.

Tema 5.- Spark Streaming

Spark Streaming es una extensión de la API principal de Spark que permite el procesamiento de flujos escalable, de alto rendimiento y tolerante a fallas de flujos de datos en tiempo real. Spark Streaming proporciona una abstracción de alto nivel llamada flujo discretizado o DStream, que representa un flujo continuo de datos.

Tema 6.- Machine Learning con MLib en Spark

MLlib es la biblioteca de aprendizaje automático (ML) de Spark. Su objetivo es hacer que el aprendizaje automático práctico sea escalable y fácil. A un alto nivel, proporciona herramientas como: Algoritmos ML: algoritmos de aprendizaje comunes como clasificación, regresión, agrupamiento y filtrado colaborativo.

5. ACTIVIDADES FORMATIVAS Y MODALIDADES DE ENSEÑANZAS

5.1 Modalidades de enseñanza

La asignatura se desarrollará a través de los siguientes métodos y técnicas generales, que se aplicarán diferencialmente según las características propias de la asignatura:

- **Método expositivo/Lección magistral:** el profesor desarrollará, mediante clases magistrales y dinámicas los contenidos recogidos en el temario.
- **Estudio de casos:** análisis de casos reales relacionados con la asignatura.
- **Resolución de ejercicios y problemas:** los estudiantes desarrollarán las soluciones adecuadas aplicando procedimientos de transformación de la información disponible y la interpretación de los resultados.
- **Aprendizaje basado en problemas:** utilización de problemas como punto de partida para la adquisición de conocimientos nuevos.
- **Aprendizaje orientado a proyectos:** se pide a los alumnos que, en pequeños grupos, planifiquen, creen y evalúen un proyecto que responda a las necesidades planteadas en una determinada situación.
- **Aprendizaje cooperativo:** Los estudiantes trabajan en grupo para realizar las tareas de manera colectiva.

5.2 Actividades formativas

Actividad Formativa	Horas	Presencialidad
AF1 Clases teóricas / Expositivas	45	100%
AF2 Clases Prácticas	36	100%
AF3 Tutorías	6	50%
AF4 Estudio independiente y trabajo autónomo del alumno	57	0%
AF5 Elaboración de trabajos (en grupo o individuales)	29	0%
AF6: Actividades de Evaluación	9	100%

6. DESARROLLO TEMPORAL

UNIDADES DIDÁCTICAS / TEMAS	PERÍODO TEMPORAL
Introducción a la asignatura, instalación Spark	Semana 1
Introducción a Python (estructuras y data frames, lambda y map/reduce/filter)	Semana 2
Spark Básico	Semanas 3 a 5
Examen Parcial 1* y Spark Cluster	Semanas 6 a 7
Spark SQL	Semanas 8 a 9
Spark Streaming	Semanas 10 a 11
Machine Learning con MLib en Spark	Semanas 12 a 13
Examen Parcial 2**, Tutoría y Presentación Proyecto Final	Semanas 14 a 15
Introducción a la asignatura, instalación Spark	Semana 1
Introducción a Python (estructuras y data frames, lambda y map/reduce/filter)	Semana 2
Spark Básico	Semanas 3 a 5
Examen Parcial 1* y Spark Cluster	Semanas 6 a 7
Spark SQL	Semanas 8 a 9
Spark Streaming	Semanas 10 a 11

* Examen Parcial 1 – contenido: Spark Básico y Python Básico

** Examen Parcial 2 – contenido: Spark SQL, Spark Streaming y Machine Learning en Spark

7. SISTEMA DE EVALUACIÓN

ACTIVIDAD DE EVALUACIÓN	VALORACIÓN MÍNIMA RESPECTO A LA CALIFICACIÓN FINAL (%)	VALORACIÓN MÁXIMA RESPECTO A LA CALIFICACIÓN FINAL (%)
SE1 Evaluación de la participación en clase, en prácticas o en proyectos de la asignatura	10%	30%
SE2 Evaluación de trabajos, proyectos, informes, memorias	40%	80%
SE3 Prueba Objetiva	10%	60%

8. CRITERIOS DE EVALUACIÓN

ACTIVIDAD DE EVALUACIÓN	VALORACIÓN MÁXIMA RESPECTO A LA CALIFICACIÓN FINAL (%)
Evaluación de trabajos, proyectos, informes, memorias	10%
Evaluación de trabajos, proyectos, informes, memorias	40%
Exámenes parciales (promedio simple)	50%
Prueba Objetiva	Reemplaza peor resultado***

- Para aprobar la asignatura en la convocatoria ordinaria, es imprescindible que la nota de las actividades/proyecto sea superior a 5.0 y la nota promedio de los exámenes parciales sea como mínimo de 5.0 (sobre 10).
- *** Caso el alumno suspenda en parciales o proyecto, tendrá la oportunidad de presentarse al examen ordinario. El examen ordinario será solo de la parte suspendida (parcial 1 o parcial 2). En caso de un suspendido en proyecto, el examen ordinario será una segunda oportunidad para presentar el proyecto de la asignatura.
- En caso de no conseguir el aprobado en la convocatoria ordinaria, el alumno podrá presentarse a la convocatoria extraordinaria de julio. En dicho caso, este examen engloba todo el contenido de la asignatura y la nota obtenida en convocatoria extraordinaria representará el 100% de la calificación del alumno en la asignatura.
- No está permitido el uso de teléfonos móviles en el aula durante el período de evaluación continua, excepto indicación expresa en sentido contrario del profesor. Los ordenadores portátiles podrán utilizarse únicamente para actividades relacionadas con la asignatura. El profesor podrá retirar el derecho al uso del ordenador a aquellos alumnos que lo utilicen para actividades que no estén relacionadas con la asignatura (consulta de correos, noticias o redes sociales, consulta o elaboración de actividades de otras asignaturas, etc.).
- No está permitido consumir bebidas ni comidas en el aula. Tampoco está permitida la presencia de cualquier tipo de bebida en las mesas, incluso en envases cerrados.
- Se demandará del alumno una participación activa, necesaria para el desarrollo de las clases.
- Se exigirá al alumno un buen comportamiento en todo momento durante el desarrollo de las clases. El mal comportamiento que impida el normal desarrollo de la clase puede conllevar la expulsión del aula por un tiempo a determinar por el profesor.

9. BIBLIOGRAFÍA / WEBGRAFÍA

Bibliografía Básica:

Holden Karau, , Learning Spark: Lightning-Fast Big Data, 2015

Bibliografía Recomendada:

Libro online de IA y Big Data: <https://iaarbook.github.io/>

10. MATERIAL, SOFTWARE Y HERRAMIENTAS NECESARIAS

TIPOLOGÍA DEL AULA:

Sala virtual del curso en Blackboard

MATERIALES DEL ALUMNO:

Ordenador personal, webcam y micrófono

SOFTWARE:

- Ubuntu directamente en el ordenador, Ubuntu en Windows o Ubuntu en Virtual Box
- Spark