



GUÍA DOCENTE

PROCESAMIENTO DE DATOS

DOBLE GRADO EN MATEMÁTICA COMPUTACIONAL E INGENIERÍA DEL SOFTWARE

MODALIDAD: PRESENCIAL

CURSO ACADÉMICO: 2023-2024

Denominación de la asignatura:	Procesamiento de Datos
Titulación:	Ingeniería del Software
Facultad o Centro:	Centro Universitario de Tecnología y Arte Digital
Materia:	Ingeniería de Datos
Curso:	3º
Cuatrimestre:	2
Carácter:	OBM
Créditos ECTS:	6
Modalidad/es de enseñanza:	Presencial
Idioma:	Castellano
Profesor/a - email	Manoel Fernando Alonso Gadi / manoel.alonso@u-tad.com
Página Web:	http://www.u-tad.com/

DESCRIPCIÓN DE LA ASIGNATURA

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 cant

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 Scala con el objetivo de utilizar Spark usando SparkShell. 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.

COMPETENCIAS Y RESULTADOS DE APRENDIZAJE

Competencias (genéricas, específicas y transversales)

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.

Resultados de aprendizaje

Al acabar la titulación, el graduado o graduada será capaz de:

- Comprender e implementar los métodos de almacenamiento y administración eficaz en entornos distribuidos de datos no estructurados.
- Conocer y saber aplicar las distintas técnicas de aprendizaje supervisado, semi-supervisado y no supervisado.
- Entender y aplicar las técnicas de Deep learning
- Ser capaz de recuperar información mediante técnicas de web scraping o APIs normalizadas
- Entender y aplicar las técnicas de análisis del lenguaje natural
- Ser capaz de analizar contenidos de redes sociales
- Entender la naturaleza y representación de las imágenes digitales.
- Conocer las aplicaciones de las redes neuronales al análisis y generación de sonido, imagen estática y video.
- Desarrollar soluciones informáticas aplicadas a la visión por computador.
- Desarrollar un proyecto completo de datos aplicando metodología iterativa, desde el diseño hasta el despliegue.

CONTENIDO

Fragmentación y distribución de datos.

Concurrencia distribuida.

Protocolos de confiabilidad y confirmación.

Administración de datos replicados.

Arquitecturas distribuidas de gestión de datos.

TEMARIO

Tema 1.- Introducción a Spark y Scala

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

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 MLlib 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.

Tema 7.- Grafos con GraphX en Spark

GraphX es una biblioteca de procesamiento de gráficos distribuida que forma parte del ecosistema Apache Spark. Esta biblioteca proporciona un conjunto de abstracciones y operaciones específicas para trabajar con datos de gráficos a gran escala de manera eficiente y distribuida

ACTIVIDADES FORMATIVAS Y METODOLOGÍAS DOCENTES

Actividades formativas

Actividad Formativa	Horas totales	Horas presenciales
<i>Clases teóricas / Expositivas</i>	29,38	29,38

<i>Clases Prácticas</i>	23,25	23,25
<i>Tutorías</i>	4,00	0,00
<i>Estudio independiente y trabajo autónomo del alumno</i>	50,00	0,00
<i>Elaboración de trabajos (en grupo o individuales)</i>	31,88	0,00
<i>Actividades de Evaluación</i>	5,25	5,25
<i>Seguimiento de Proyectos</i>	6,25	6,25
TOTAL	150	64,13

Metodologías docentes

Método expositivo o lección magistral

Aprendizaje de casos

Aprendizaje basado en la resolución de problemas

Aprendizaje basado en proyectos

Aprendizaje cooperativo o colaborativo

Aprendizaje por indagación

Metodología Flipped classroom o aula invertida

Gamificación

Just in time Teaching (JITT) o aula a tiempo

Método expositivo o lección magistral

Método del caso

Aprendizaje basado en la resolución de problemas

Aprendizaje basado en proyectos

Aprendizaje cooperativo o colaborativo

Aprendizaje por indagación

Metodología flipped classroom o aula invertida

Gamificación

DESARROLLO TEMPORAL

UNIDADES DIDÁCTICAS / TEMAS PERÍODO TEMPORAL

Introducción a la asignatura, instalación de máquina virtual Semana 1

Introducción a scala (estructuras y funciones, lambda y map/reduce/filter) Semana 2

Spark Básico Semanas 3 a 5

Spark Cluster Semanas 6 a 7

Spark SQL Semanas 8 a 9

Parcial Semanas 10

Spark Streaming Semanas 11 a 12

Machine Learning con MLib en Spark Semanas 13 a 14

Spark GraphX Semanas 15

SISTEMA DE EVALUACIÓN

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

CRITERIOS ESPECÍFICOS DE EVALUACIÓN

ACTIVIDAD DE EVALUACIÓN	CONVOCATORIA ORDINARIA	CONVOCATORIA EXTRAORDINARIA
<i>Evaluación de la participación en clase, en prácticas o en proyectos de la asignatura</i>	10	10
<i>Evaluación de trabajos, proyectos, informes, memorias</i>	40	40

<i>Prueba Objetiva</i>	50	50
------------------------	----	----

Consideraciones generales acerca de la evaluación

- La nota final de la asignatura en convocatoria ordinaria se calculará con las notas de las tres actividades descritas en la tabla anterior aplicándoles los pesos correspondientes. Por tanto, la regla a aplicar para calcular la nota final será:

$$\text{Nota_Final_Ordinaria} = 10\% \cdot \text{AE1} + 40\% \cdot \text{AE2} + 50\% \cdot \text{AE3}$$

- AE1: La evaluación de la participación y el trabajo de clase: se realizará a partir de la asistencia, el trabajo en clase y la entrega de las actividades y ejercicios propuestos durante las clases. Este aspecto representará el 10% de la calificación final de la asignatura en la convocatoria ordinaria.

- AE2: A lo largo del curso se plantearán práctica(s) obligatoria(s) que deberán ser entregada(s) en la fecha indicada a través de la plataforma virtual. Adicionalmente, se realizará además un examen parcial de la materia impartida hasta el momento. Esta actividad (práctica(s) obligatoria(s) y parcial) se evaluarán a través de la propia plataforma virtual y supondrá un 40% de la calificación final de la asignatura en la convocatoria ordinaria. No se admitirán trabajos fuera de forma y fecha sin causa justificada y todos ellos son de presentación obligatoria para aprobar en convocatoria ordinaria. Si alguna de estas prácticas evaluables obligatorias se entrega fuera de plazo tendrá una penalización en la calificación del 30%. Para poder aprobar la asignatura en convocatoria ordinaria será requisito haber entregado la(s) práctica(s) obligatoria(s) que se hayan solicitado. Si el alumno no ha realizado el parcial propuesto tendrá un 0 en dicha calificación.

- AE3: Al final del cuatrimestre se realizará un examen final, que tendrá un peso total sobre la nota de la convocatoria ordinaria de un 50%. Para aprobar la asignatura en convocatoria ordinaria será necesario que el alumno tenga en dicho examen al menos un 4 (sobre 10).

- Resumiendo, para aprobar la asignatura en convocatoria ordinaria es necesario que el examen final sea al menos de 4.0 (sobre 10), que estén entregadas la(s) práctica(s) evaluables solicitadas durante el curso (las pertenecientes a la actividad AE2 del 40%), es decir, dichas prácticas son obligatorias y no se hará media si alguna de ellas no está presentada y que la nota media final (incluyendo AE1, AE2 y AE3) sea al menos de 5.0 (sobre 10).

En caso de no cumplirse alguno de estos requisitos, la asignatura se considerará automáticamente suspensa independientemente del resto de calificaciones.

- En caso de no conseguir el aprobado en la convocatoria ordinaria, el alumno podrá presentarse a la convocatoria extraordinaria. A saber:

- Si el alumno tiene la materia suspensa en convocatoria ordinaria por no haber alcanzado el 4 en el examen final, pero tiene todas sus prácticas entregadas y aprobadas, entonces se tendrá que presentar

exclusivamente al examen extraordinario y se le guardan el resto de sus notas del curso. Su examen extraordinario mantiene el peso del 50% de la nota y debe tener una calificación mínima de 4 (sobre 10).

- Si el alumno tiene la materia suspensa en ordinaria por prácticas pendientes y/o suspensas pero su nota del examen es mayor o igual que 4, se le guarda dicho examen, pero tendrá que presentar las prácticas pendientes o suspensas en convocatoria extraordinaria.
- En caso contrario, (examen suspenso y prácticas obligatorias suspensas) deberá:
- Realizar el examen final extraordinario que representará el 60% de la calificación en dicha convocatoria y en el que la materia exigible al alumno será todo el contenido de la asignatura visto en clase.
- Además, deberá entregar la(s) práctica(s) suspensas o no entregadas y que se planteen al comienzo del segundo cuatrimestre y cuyo peso sobre la nota final será el 30%.
- El 10% restante será la nota de la asistencia y trabajo en clase obtenida durante el curso.

La nota media final debe ser al menos de 5.0 (sobre 10) y la nota del examen final debe ser al menos un 4 (sobre 10).

- No se conservarán calificaciones de ningún tipo entre distintos cursos académicos, ni entre distintas convocatorias.

Detalles de Evaluación Ordinaria

Como apruebo la convocatoria ordinaria:

2. Para aprobar en la convocatoria ordinaria el alumno deberá tener una nota igual o superior a 5,00 en la media final de todas las calificaciones (incluyendo la nota de práctica(s) y parcial, la nota del examen final y la nota de asistencia y el trabajo en clase).
3. Para que pueda realizarse dicha media el alumno deberá tener al menos un 4,00 en el Examen Final y haber entregado la(s) práctica(s) obligatorias solicitadas durante el curso.
4. Para que las actividades prácticas obligatorias evaluables sean calificadas será necesario haber realizado las entregas en la fecha establecida por el profesor, es decir, no se admitirán trabajos fuera de forma y fecha sin causa justificada. Los retrasos injustificados de las entregas tendrán una penalización de la nota del 30%.

Detalles Evaluación Extraordinaria

Cuando debo hacer la convocatoria extraordinaria y sus condiciones:

5. Los alumnos con menos de 4 en el examen final ordinario pero con las actividades prácticas obligatorias (las del grupo AE2) presentadas y aprobadas deberán presentarse exclusivamente al examen de convocatoria extraordinaria y se le guardan el resto de notas del curso. Su nota media de la convocatoria extraordinaria será por tanto:

Nota final convocatoria extraordinaria = 10% asistencia y trabajo de clase

+ 40% de práctica(s) y parcial + 50% Examen extraordinario

- Si el alumno tiene la materia suspensa en ordinaria por prácticas obligatorias pendientes y/o suspensas pero su nota del examen es mayor o igual que 4, se le guarda dicho examen, pero tendría que presentar la(s) práctica(s) pendientes o suspensas en convocatoria extraordinaria.

Nota final convocatoria extraordinaria = 10% asistencia y trabajo de clase

+ 40% de práctica(s) conv. extraordinaria + 50% Examen Ordinaria

6. El resto de alumnos con la materia suspensa deberán presentarse a un examen final extraordinario y entregar en extraordinaria la(s) práctica(s) que haya definido y solicitado el profesor para su realización durante el Q2, siendo su nota final:

Nota final convocatoria extraordinaria = 10% asistencia y trabajo de clase

+ 30 % practica(s) conv. extraordinaria + 60% Nota del Examen extraordinario

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/>

MATERIALES, SOFTWARE Y HERRAMIENTAS NECESARIAS

Tipología del aula

Aula teórica

Equipo de proyección y pizarra

Materiales:

Ordenador personal

Software:

Virtual Box

Ubuntu 22.04