



GUÍA DOCENTE CURSO: 2019-20

DATOS BÁSICOS DE LA ASIGNATURA

Asignatura:	Computación sobre Datos Masivos		
Código de asignatura:	71144211	Plan:	Máster en Tecnologías y Aplicaciones en Ingeniería Informática
Año académico:	2019-20	Ciclo formativo:	Máster Universitario Oficial
Curso de la Titulación:	1	Tipo:	Optativa
Duración:			

DISTRIBUCIÓN HORARIA DE LA ASIGNATURA SEGÚN NORMATIVA

Créditos:	4
Horas totales de la asignatura:	100
UTILIZACIÓN DE LA PLATAFORMA VIRTUAL:	Multimodal

DATOS DEL PROFESORADO

Nombre	Martínez Masegosa, Irene		
Departamento	Dpto. de Informática		
Edificio	Edificio Científico Técnico III Matemáticas e Informática (CITE III). Planta 2		
Despacho	620		
Teléfono	+34 950 015679	E-mail (institucional)	irene@ual.es
Recursos Web personales	http://cms.ual.es/UAL/personas/persona.htm?id=505553505157495275		

ELEMENTOS DE INTERÉS PARA EL APRENDIZAJE DE LA ASIGNATURA

Justificación de los contenidos

La importancia del Big Data descansa en la inmensa cantidad de datos que se generan cada día, especialmente a raíz del crecimiento de las redes sociales, del número de dispositivos (smartphones, wearables, etc) o redes de sensores, así como en la posibilidad de tener en cuenta información actualizada y variada para la toma de decisiones. Google no existiría sin los datos masivos, tampoco Facebook.

Las mismas organizaciones que recopilan grandes volúmenes de datos suelen tener dificultades a la hora de utilizarlos, por una serie de razones: la arquitectura de los productos analíticos tradicionales no está pensada para la computación distribuida, y los algoritmos estadísticos existentes no están diseñados para trabajar con cantidades masivas de datos.

Por tanto, el análisis de datos masivos con tecnología puntera requiere nuevas habilidades y una búsqueda de soluciones que permitan escalar bien a datos realmente masivos mediante arquitecturas distribuidas centradas en datos.

Materia con la que se relaciona en el Plan de Estudios

Esta asignatura se encuentra dentro del **módulo de Big Data** y se relaciona estrechamente con el resto de las asignaturas de este módulo.

Conocimientos necesarios para abordar la Asignatura

Se considera que el alumno que puede matricularse en este máster ha obtenido los conocimientos previos necesarios para abordar esta asignatura.

Requisitos previos recogidos en la memoria de la Titulación

Los propios de acceso a la titulación de Máster.

COMPETENCIAS

Competencias Básicas y Generales

Competencias Básicas

- Aplicación de conocimientos

Competencias Transversales de la Universidad de Almería

- Capacidad para resolver problemas

Competencias Específicas desarrolladas

CE04 - Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería Informática.

TI06 - Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.

TI07 - Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería

OBJETIVOS/RESULTADOS DEL APRENDIZAJE

Que el estudiante sea capaz de aplicar los conocimientos adquiridos y resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con la computación de datos masivos.

Que el estudiante sea capaz de llevar a cabo el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática y en particular en el campo de los datos masivos o big data.

Que el estudiante sea capaz de resolver problemas derivados del uso y explotación de datos masivos.

Que el estudiante sea capaz de diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.

Que el estudiante sea capaz de comprender y aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería relacionados con el uso de datos masivos.

PLANIFICACIÓN

Temario

1. Computación sobre datos masivos
 1. Características
 2. Patrones básicos de diseño
 3. Sistemas de almacenamiento
 4. Modelos de procesamiento distribuido
2. Herramientas para el procesamiento distribuido de datos masivos
 1. *Frameworks* y estándares básicos
 2. Procesamiento batch y procesamiento en streaming
3. Aprendizaje automático y optimización
 1. Herramientas para el análisis de datos en tiempo real
 2. Herramientas de aprendizaje automático

Metodología y Actividades Formativas

Clases magistrales teórico-prácticas participativas

Resolución de problemas

Trabajo en grupo

Elaboración y redacción de trabajos prácticos

Presentación de trabajos/proyectos

Tutorías

Actividades de Innovación Docente

Diversidad Funcional

Aquellos estudiantes con discapacidad o necesidades educativas especiales pueden dirigirse a la Delegación del Rector para la Diversidad Funcional (<http://www.ual.es/discapacidad>) para recibir la orientación o asesoramiento oportunos y facilitar un mejor aprovechamiento de su proceso formativo. De igual forma podrán solicitar la puesta en marcha de las adaptaciones de contenidos, metodología y evaluación necesarias que garanticen la igualdad de oportunidades en su desarrollo académico. El tratamiento de la información sobre este alumnado, en cumplimiento con la LOPD, es de estricta confidencialidad. Los docentes responsables de esta guía aplicaran las adaptaciones aprobadas por la Delegación, tras su notificación al Centro y al coordinador de curso

PROCEDIMIENTO DE EVALUACIÓN DE LAS COMPETENCIAS

Criterios e Instrumentos de Evaluación

Los contenidos se estudian dentro del marco de dos Frameworks. Cada uno de ellos tiene uno o varios trabajos asociados.

Los trabajos se pueden realizar de forma individual o en grupo, dependiendo del tema. Es necesario obtener una puntuación mínima en los trabajos/pruebas de cada uno de los Frameworks.

Los porcentajes de la nota final son los siguientes:

- Evaluación de trabajos presentados mediante pruebas orales/escritas: 70%
- Asistencia, defensa de trabajos y participación activa en clase: 20%
- Acceso y participación en el aula virtual: 10%

Los estudiantes que no asistan a clase podrán obtener como máximo una nota final correspondiente al 70% por evaluación de trabajos/pruebas, siendo necesaria una calificación mínima de 5 puntos para aprobar la asignatura.

Las competencias a evaluar de la asignatura son:

- CE04 - Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería Informática.
- CB7 - 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.
- CT01 - Capacidad para resolver problemas.
- TI06- Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida, y
- TI07- Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería

La evaluación de trabajos presentados mediante pruebas orales/escritas: evalúa las competencias CE04, CB7, CT01, TI06, TI07. La asistencia y participación activa en clase: evalúa las competencias CE04, CB7 CT01, TI06 y TI07. El acceso y participación en el aula virtual : CT01 y CE04

Mecanismos de seguimiento

- Asistencia y participación en seminarios
- Alta y acceso al aula virtual
- Participación en herramientas de comunicación (foros de debate, correos)
- Entrega de actividades en clase
- Entrega de actividades en aula virtual

BIBLIOGRAFÍA

Bibliografía recomendada

Básica

- Tom White. Hadoop: the definitive guide. O'Reilly. 2015.
- Boris Lublinsky, Kevin T.Smith, Alexey Yakubovich. Professional Hadoop Solutions. John Wiley & Sons. 2013.
- Mauro Macías, Mauro Gómez. Introduccion a Apache Spark. Editorial UOC. 2015.
- Holden Karau; Andy Konwinski; Patrick Wendell; Matei Zaharia. Learning Spark : lightening fast data analysis. O'Reilly Media. 2015.
- Nathan Marz ; James Warren. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications. 2015.

Complementaria

- Vijay Srinivas Agneeswaran. Big data analytics beyond Hadoop: : real-time applications with Storm, Spark, and more Hadoop alternatives. Pearson Education. 2014.
- Juan Carlos González Cabañas, José Francisco Aldana Montes, Ismael Navas Delgado. Introduccion al big data. García Maroto Editores. 2016.

Otra Bibliografía

Bibliografía existente en el Sistema de Información de la Biblioteca de la UAL

Puede ver la bibliografía existente en la actualidad en el Sistema de Gestión de Biblioteca consultando en la siguiente dirección:

[http://almirez.ual.es/search/e?SEARCH=COMPUTACION SOBRE DATOS MASIVOS](http://almirez.ual.es/search/e?SEARCH=COMPUTACION+SOBRE+DATOS+MASIVOS)

DIRECCIONES WEB

- <http://hadoop.apache.org/>
web Apache Hadoop
- <https://spark.apache.org/>
web Apache Spark