

Guía docente de la asignatura

## Big Data II

Fecha última actualización: 18/07/2021  
Fecha de aprobación por la Comisión Académica: 26/07/2021

**Máster**

Máster Universitario en Ciencia de Datos e Ingeniería de Computadores

**MÓDULO**

Módulo Big Data y Cloud Computing

**RAMA**

Ingeniería y Arquitectura

**CENTRO RESPONSABLE DEL TÍTULO**

Escuela Internacional de Posgrado

**Semestre**

Segundo

**Créditos**

4

**Tipo**

Optativa

**Tipo de enseñanza**

Presencial

### BREVE DESCRIPCIÓN DE CONTENIDOS (Según memoria de verificación del Máster)

Big data, procesamiento de datos masivos, analítica, preprocesamiento y aprendizaje de datos con big data.

### COMPETENCIAS

#### COMPETENCIAS BÁSICAS

- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- 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.
- CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- 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.



### COMPETENCIAS GENERALES

- CG01 - Capacidad de acceso y gestión de la información
- CG02 - Capacidad de análisis y síntesis
- CG03 - Capacidad de organización y planificación
- CG04 - Capacidad emprendedora
- CG05 - Capacidad para tomar decisiones de forma autónoma
- CG07 - Motivación por la calidad
- CG08 - Capacidad para trabajar en equipo

### COMPETENCIAS ESPECÍFICAS

- CE03 - Capacidad para la aplicación de técnicas y metodologías que permitan abordar desde nuevas perspectivas los problemas de interés, gracias a la disponibilidad de las plataformas de computación y comunicación con altos niveles de prestaciones.
- CE04 - Capacidad de análisis de aplicaciones en ámbitos de biomedicina y bioinformática, optimización y predicción, control avanzado, y robótica bioinspirada, tanto desde el punto de vista de los requisitos para una implementación eficaz de los algoritmos y las técnicas de computación que se usan para abordarlas, como de las características deseables en las arquitecturas donde se ejecutan
- CE05 - Capacidad para modelar y resolver problemas reales o académicos mediante técnicas de ciencia de datos
- CE06 - Capacidad para modelar y resolver problemas reales o académicos mediante tecnologías inteligentes o de inteligencia computacional.
- CE08 - Conocer algunas de las principales aplicaciones de la ciencia de datos y de las tecnologías inteligentes

### COMPETENCIAS TRANSVERSALES

- CT01 - Ser consciente de la importancia del desarrollo sostenible y demostrar sensibilidad medioambiental.
- CT02 - Ser consciente del derecho a la no discriminación y al acceso universal al conocimiento de las personas con discapacidad.

### RESULTADOS DE APRENDIZAJE (Objetivos)

El estudiante o la estudiante aprenderá a manejar las tecnologías de cloud computing, y almacenamiento y procesamiento de grandes volúmenes de datos bajo las tecnologías actuales de Big Data, tales como Hadoop y Spark.

En particular, el estudiante o la estudiante:

- Aprenderá cómo resolver problemas de aprendizaje en Big Data, aplicando las distintas técnicas conocidas para la etapa de modelado, así como técnicas de preprocesamiento de datos en aquellos problemas en los que la etapa de preprocesamiento sea necesaria o recomendable, todo ello bajo las tecnologías distribuidas de cloud computing y big data.

Al mismo tiempo y de manera práctica, el estudiante o la estudiante deberá:

- Saber aplicar las técnicas anteriormente mencionadas para modelar y resolver problemas reales o académicos de big data.



- Conocer herramientas software existentes para el manejo de información en Big Data, como el lenguaje Pig Latin y la librería MLLIB, junto con el repositorio Spark Package.

## PROGRAMA DE CONTENIDOS TEÓRICOS Y PRÁCTICOS

### TEÓRICO

Temario teórico (temas):

1. Programación y procesamiento de datos. Lenguaje Pig Latin y software Apache Hive
2. Big Data Analytics: Diseño de algoritmos bajo el paradigma MapReduce: Aprendizaje supervisado y no supervisado
3. Smart Data. Preprocesamiento de datos en Big Data

### PRÁCTICO

Temario práctico (seminarios):

- Programación y procesamiento de datos con Pig Latin
- Plataforma para la analítica de datos: Spark
- Bibliotecas de Algoritmos: MLLib, ...

Prácticas de Laboratorio:

1. Caso de estudio con Pig Latin
2. Uso de la librería MLLIB con Spark. Casi practia
3. Uso de los algoritmos del repositorio Spark Package.

## BIBLIOGRAFÍA

### BIBLIOGRAFÍA FUNDAMENTAL

- Mining of Massive Datasets, Anand Rajaraman, Jure Leskovec, Jeffrey D. Ullman, Cambridge University Press, 2011
- Machine Learning with Spark, Nick Pentreath, PACKT Publishing, 2015.
- Advanced Analytics with Spark, Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, O'Reilly, 2015.
- Big Data Preprocessing. Enabling Smart Data, J. Luengo, D. García-Gil, S. Ramírez-Gallego, S. García López, F. Herrera Springer, 2020

### BIBLIOGRAFÍA COMPLEMENTARIA

## ENLACES RECOMENDADOS



- Sitio web del Máster Universitario Oficial en Ciencia de Datos e Ingeniería de Computadores: <https://masteres.ugr.es/datcom/>
- Fundación Apache: <https://www.apache.org/>

## METODOLOGÍA DOCENTE

- MD01 Lección magistral/expositiva
- MD02 Resolución de problemas y estudio de casos prácticos
- MD03 Prácticas de laboratorio
- MD04 Seminarios
- MD05 Análisis de fuentes y documentos
- MD06 Realización de trabajos en grupo
- MD07 Realización de trabajos individuales
- MD08 Seguimiento del TFM

## EVALUACIÓN (instrumentos de evaluación, criterios de evaluación y porcentaje sobre la calificación final)

### EVALUACIÓN ORDINARIA

Se realizará mediante una prueba en clase que corresponderá al 20% de la calificación y un trabajo práctico que corresponde al 80%.

Para el trabajo práctico, se le entregará a cada estudiante un conjunto de bases de datos y se pedirá que aplique las herramientas estudiadas en Pig, y Spark.

- (25%) Construir un script en Pig Latin para el procesamiento de datos en big data
- (55%) Dar un modelo para un problema de aprendizaje utilizando la biblioteca MLlib (Spark), ... Hay que describir detalladamente todo el proceso algorítmico utilizado, mostrando los resultados de cada uno de los algoritmos utilizados para entrenamiento y test.

Se habilitará una única entrega de actividad en la que se entregará un fichero comprimido que incluirá un fichero de solución o respuesta para los enunciados. Es obligatorio entregar todas las actividades para obtener la calificación. La calificación final estará asociada a la correcta resolución de los enunciados.

### EVALUACIÓN EXTRAORDINARIA

La evaluación se llevará a cabo mediante la realización de un trabajo que incluirá:

- (25%) la construcción de un script en Pig Latin para el procesamiento de datos en Big Data
- (75%) la propuesta de un modelo para un problema de aprendizaje utilizando la biblioteca MLlib (Spark), ... incluyendo la descripción detallada de todo el proceso algorítmico utilizado, mostrando los resultados de cada uno de los algoritmos utilizados para entrenamiento y test.

Se habilitará una única entrega de actividad en la que se entregará un fichero comprimido que



incluirá un fichero de solución o respuesta para cada uno de los cuatro enunciados. Es obligatorio entregar todas las actividades para obtener la calificación. La calificación final estará asociada a la correcta resolución de los cuatro enunciados.

### EVALUACIÓN ÚNICA FINAL

La evaluación se llevará a cabo mediante la realización de un trabajo que incluirá:

- (25%) la construcción de un script en Pig Latin para el procesamiento de datos en Big Data
- (75%) la propuesta de un modelo para un problema de aprendizaje utilizando la biblioteca MLLib (Spark), ... incluyendo la descripción detallada de todo el proceso algorítmico utilizado, mostrando los resultados de cada uno de los algoritmos utilizados para entrenamiento y test.

Se habilitará una única entrega de actividad en la que se entregará un fichero comprimido que incluirá un fichero de solución o respuesta para cada uno de los cuatro enunciados. Es obligatorio entregar todas las actividades para obtener la calificación. La calificación final estará asociada a la correcta resolución de los cuatro enunciados.

