

Guía docente de la asignatura

Big Data II (M51/56/3/36)

Fecha de aprobación por la Comisión Académica: 26/07/2023

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

PRERREQUISITOS Y/O RECOMENDACIONES

Se recomienda haber cursado la asignatura de Big Data I.

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

Big Data, procesamiento de datos masivos, analítica, pre-procesamiento 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 pre-procesamiento de datos en aquellos problemas en los que dicha etapa 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 los lenguajes de proceso de datos masivos por lotes y la biblioteca MLLIB, junto con el repositorio Spark Package.

PROGRAMA DE CONTENIDOS TEÓRICOS Y PRÁCTICOS

TEÓRICO

Temario teórico (temas):

1. Programación y procesamiento de datos. Lenguaje de procesamiento de datos masivo por lotes.
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.
4. Procesamiento distribuido, fusión de modelos, y privacidad de datos: Aprendizaje Federado.

PRÁCTICO

Temario práctico (seminarios):

- Programación y procesamiento de datos con un lenguaje de procesamiento de datos masivos por lotes.
- Plataformas para la analítica de datos. Un caso: Spark.
- Biblioteca de algoritmos de aprendizaje automático (MLlib).

Prácticas de Laboratorio:

1. Caso de estudio con un lenguaje de procesamiento de datos masivos por lotes.
2. Uso de la librería MLLIB con Spark. Caso práctico.
3. Uso de los algoritmos de Spark Packages y otros repositorios.

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.
- Large-scale Data Analytics with Python and Spark: A Hands-on Guide to Implementing Machine Learning Solutions, I. Triguero, M. Galar, Cambridge University Press, 2023.

BIBLIOGRAFÍA COMPLEMENTARIA

- The Apache Software Foundation. "Spark Python API: PySpark Documentation" (2022). Disponible en: <https://spark.apache.org/docs/latest/api/python/index.html>
- The Apache Software Foundation. "MLlib" (2022). Disponible en: <https://spark.apache.org/docs/latest/ml-guide.html>
- The Apache Software Foundation. "Spark Scala API" (2022). Disponible en: <https://spark.apache.org/docs/latest/api/scala/org/apache/spark/index.html>

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

La evaluación de conocimientos adquiridos en la Convocatoria Ordinaria constará de los siguientes ítems de evaluación:

- Pre-procesamiento de datos mediante lenguajes de procesamiento por lotes:
 1. (25%) construir un script en el lenguaje de procesamiento de datos masivos para el pre-procesamiento en big data
- Analítica de datos:
 2. (20%) prueba escrita sobre analítica de datos realizada en horario de clase, y
 3. (55%) diseñar un modelo para un problema de aprendizaje utilizando la



biblioteca MLLib (Spark), ... en el que habrá que describir detalladamente todo el proceso algorítmico utilizado, mostrando los resultados de cada uno de los algoritmos utilizados para entrenamiento y test

Para los ítems de evaluación 1 y 3 se proporcionarán sendas actividades en la Plataforma Docente en las que se proporcionarán los detalles del trabajo solicitado y se fijarán las fechas para la entrega de dichos trabajos. Ambas actividades estarán a disposición del estudiantado una vez concluidas las sesiones dedicadas a la parte evaluada por el correspondiente trabajo.

Para superar la asignatura en la Convocatoria Ordinaria, será necesario haber obtenido calificación en dos o más ítems de evaluación y que la suma ponderada de las calificaciones obtenidas iguale o supere la calificación de 5 sobre 10.

EVALUACIÓN EXTRAORDINARIA

La evaluación de conocimientos adquiridos en la Convocatoria Extraordinaria constará de los siguientes ítems de evaluación:

1. (25%) Pre-procesamiento de datos mediante lenguajes de procesamiento por lotes: construir un script en el lenguaje de procesamiento de datos masivos para el pre-procesamiento en Big Data
2. (75%) Analítica de datos: diseñar un modelo para un problema de aprendizaje utilizando la biblioteca MLLib (Spark), ... en el que habrá que describir detalladamente todo el proceso algorítmico utilizado, mostrando los resultados de cada uno de los algoritmos utilizados para entrenamiento y test

Si el ítem de evaluación de conocimientos y habilidades en el "pre-procesamiento de datos mediante lenguajes de procesamiento por lotes" (correspondiente al 25% de la calificación de la asignatura) hubiera sido evaluado en la Convocatoria Ordinaria con una calificación igual o superior a 5 sobre 10, el estudiante o la estudiante podrá optar por no evaluar dicho ítem en la Convocatoria Extraordinaria, conservando la calificación obtenida en la Convocatoria Ordinaria en dicho ítem.

Para someter a evaluación ambos ítems, se proporcionarán sendas actividades en la Plataforma Docente en las que se proporcionarán los detalles del trabajo solicitado y se fijarán las fechas para la entrega de dichos trabajos.

Para superar la asignatura en la Convocatoria Extraordinaria, será necesario que la suma ponderada de las calificaciones obtenidas supere la calificación de 5 sobre 10.

EVALUACIÓN ÚNICA FINAL

La evaluación de conocimientos adquiridos a través del procedimiento de Evaluación Única Final constará de los siguientes ítems de evaluación:

1. (25%) Pre-procesamiento de datos mediante lenguajes de procesamiento por lotes: construir un script en el lenguaje de procesamiento de datos masivos para el pre-procesamiento en Big Data
2. (75%) Analítica de datos: diseñar un modelo para un problema de aprendizaje utilizando la biblioteca MLLib (Spark), ... en el que habrá que describir detalladamente todo el proceso algorítmico utilizado, mostrando los resultados de cada uno de los algoritmos utilizados para entrenamiento y test



Para someter a evaluación ambos ítems, se proporcionarán sendas actividades en la Plataforma Docente en las que se proporcionarán los detalles del trabajo solicitado y se fijarán las fechas para la entrega de dichos trabajos.

Para superar la asignatura a través de Evaluación Única Final, será necesario que la suma ponderada de las calificaciones obtenidas supere la calificación de 5 sobre 10.

INFORMACIÓN ADICIONAL

Información de interés para estudiantado con discapacidad y/o Necesidades Específicas de Apoyo Educativo (NEAE): [Gestión de servicios y apoyos](https://ve.ugr.es/servicios/atencion-social/estudiantes-con-discapacidad) (<https://ve.ugr.es/servicios/atencion-social/estudiantes-con-discapacidad>).

