Herramientas Informáticas para Big Data

Curso 2023-24

Módulo: Especialización
Materia: Tecnologías de la Información y la Decisión
Créditos: 3 ECTS

Contenido:

1. Introducción al Big Data.
2. Almacenamiento de datos en ambientes Big Data.
3. Sistemas de Ficheros Distribuidos Hadoop (HDFS).
4. Técnicas de procesamiento paralelo: Hadoop Map Reduce, Spark.
5. Módulos avanzados de Spark.

Resultados del aprendizaje:

- Comprensión del marco conceptual del Big Data.
- Entendimiento y manejo básico de un sistema HDFS Hadoop.
- Diseño de soluciones paralelizables utilizando el esquema Map Reduce.
- Programación en Spark.
- Conocimiento y uso de la abstracción de datos RDD y Dataframe.
- Conocimiento de algunos módulos avanzados de Spark: Streaming, SQL, ML…
- Utilización de entornos interactivos para programar y analizar datos.

Competencias:

Básicas y generales: CG1, CG6

Transversales: CT1, CT2

Específicas: CE1, CE2, CE3, CE4, CE5, CE6, CE7

Bibliografía:

A. BOSCHETTI & L. MASSARON (2015). *Python Data Science Essentials*. Birmingham: Packt Publishing.

J. VANDERPLAS (2016). *Python Data Science Handbook*. Sebastopol, CA: O'Reilly.

APACHE SOFTWARE FOUNDATION (2008-2023). *Apache Hadoop Documentation (3.3.6)*. Online, url: https://hadoop.apache.org/docs/current/

YELP AND CONTRIBUTORS (2009-2018). *Mrjob Documentation (v0.7.4)*. Online, url: https://mrjob.readthedocs.io/en/latest/

APACHE SOFTWARE FOUNDATION (2008-2023). *Apache Spark Documentation (3.5.0)*. Online, url: https://spark.apache.org/docs/latest/index.html

H. KARAU, A. KONWINSKI, P. WENDELL & M. ZAHARIA (2015). *Learning Spark: Lightning-Fast Data Analysis*. Sebastopol, CA: O'Reilly.

Metodologías docentes:

Método expositivo.
Estudio de casos.
Prácticas de ordenador.

Evaluación:

Entrega de trabajos: 40%
Examen teórico-práctico: 40%
Asistencia y participación activa: 20%

El máster es presencial y la asistencia es obligatoria.

En la convocatoria ordinaria la evaluación es continua y la entrega de prácticas se hará a medida que avance el curso.

El examen de la convocatoria extraordinaria se realizará el día de la fecha fijada en el calendario de exámenes.

Profesorado:

Nombre: Carlos Gregorio Rodríguez
Despacho: 451, Facultad de CC Matemáticas, UCM
E-mail: cgr@sip.ucm.es
Teléfono: +34 91 3944527