

Herramientas Informáticas para Big Data

Créditos: 3 ECTS

Estructura

Módulo : Especialización

Materia: Tecnologías de la Información y la Decisión

Contenido:

- Introducción a Big Data.
- Almacenamiento de datos en ambientes Big Data.
- Sistemas de Ficheros distribuidos Hadoop HDFS.
- Técnicas de procesamiento paralelo:
 - Hadoop Map Reduce
 - Spark
- Módulos avanzados

Resultados del aprendizaje:

- Comprender el marco conceptual del BigData
- Entender y manejar básicamente un sistema HDFS Hadoop
- Diseñar soluciones paralelizables utilizando el esquema Map Reduce
- Escribir programas en Spark
- Conocer la abstracción de datos RDD
- Conocer los módulos avanzados de Spark: Streaming, SQL, ML...
- Utilización de entornos interactivos para programar y analizar datos.

Competencias:

- Aprender a aplicar los conocimientos adquiridos y a explotar su potencial para la resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) en el tratamiento estadístico computacional de la información.
- Saber aplicar sus conocimientos a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y en la resolución de problemas y estudio de casos. Esto implica, más concretamente: Integrar creativamente conocimientos y aplicarlos a la resolución de problemas complejos, perseguir objetivos de calidad en el desarrollo de su actividad profesional, adquirir capacidad para la toma de decisiones y de dirección de recursos humanos, ser capaz de mostrar creatividad, iniciativa y espíritu emprendedor para afrontar los retos de su actividad, valorar la importancia de los métodos estadístico-computacionales en el contexto industrial, económico, administrativo, medio ambiental y social.
- Tener la capacidad de reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole científica, tecnológica y

empresarial. Demostrar razonamiento crítico y gestionar información científica y técnica de calidad, bibliografía, bases de datos especializadas y recursos accesibles a través de Internet.

- Adquisición de una formación sólida y rigurosa en temas avanzados de Estadística Matemática y Tecnologías de la Decisión aplicadas al tratamiento de la Información.
- Capacidad para planificar la resolución de un problema en función de las herramientas de que se disponga y, en su caso, de las restricciones de tiempo y recursos.
- Capacidad para utilizar aplicaciones informáticas estadísticas, de cálculo numérico y simbólico, visualización gráfica, optimización u otras para resolver problemas con un elevado grado de complejidad.
- Desarrollar habilidades de aprendizaje en Estadística Computacional y Matemáticas, así como en sus respectivas aplicaciones, que permitan al alumno continuar estudiando y profundizando en la materia de modo autónomo, así como el desarrollo profesional con un alto grado de independencia.
- Resolver problemas y casos reales planteados en el tratamiento estadístico computacional de la información generada en los ámbitos de la ciencia, la tecnología y la sociedad mediante habilidades de modelización matemática, estimación y computación.
- Desarrollar programas que resuelvan problemas matemáticos utilizando para cada caso el entorno computacional adecuado.
- Capacidad de utilización de herramientas de búsqueda de recursos bibliográficos así como manejo, gestión y análisis de grandes bases de datos.

Bibliografía:

- L. Massaron; A. Boschetti, Python Data Science Essentials, Packt. 2015
- J. VanderPlas, Python Data Science Handbook, O'Reilly. 2016
- Documentación Hadoop: <http://hadoop.apache.org/docs/current/>
- Documentación MrJob: <https://pythonhosted.org/mrjob/>
- Documentación Spark: <http://spark.apache.org/docs/latest/>
- Karau H., Konwinski A., Wendell P., and Zaharia M. Learning Spark. O'Reilly. 2015

Metodología docente

- Método expositivo
- Estudio de casos
- Prácticas de ordenador

Evaluación:

Para la convocatoria ordinaria, la evaluación de la asignatura es continua.

Los factores que se valoran para la calificación final son:

- Asistencia y participación. 20% (Se recuerda que el máster es presencial y la asistencia es obligatoria)
- Entrega de prácticas: 40% (a medida que avanza el curso)
- Examen teórico-práctico: 40%

Para la convocatoria extraordinaria, se valorarán los siguientes factores

- Asistencia y participación. 20% (Se recuerda que el máster es presencial y la asistencia es obligatoria)

- Examen teórico-práctico: 80% (A realizar el día de la fecha fijada en la convocatoria extraordinaria)

Profesor:

- Carlos Gregorio Rodríguez (Python)
- Despacho 451
- Facultad de Ciencias Matemáticas
- Teléfono: 913 944 527
- Correo electrónico: cgr@sip.ucm.es