



Proyecto docente

Asignatura	Modelos de Programación para el Big Data		
Materia	Tecnologías Informáticas para el Big Data		
Titulación	Máster Universitario en Inteligencia de Negocio y Big Data en Entornos Seguros		
Plan		Código	
Periodo de impartición	Primer semestre	Tipo/Carácter	Obligatoria
Nivel/Ciclo	Máster	Curso	1
Créditos ECTS	3		
Lengua en que se imparte	Castellano		
Profesor/es responsable/s	Álvar Arnaiz González		
Datos de contacto (e-mail, teléfono...)	alvarag@ubu.es 947 25 93 58		
Horario de tutorías			
Coordinador	Álvar Arnaiz González		
Departamento	Área de Lenguajes y Sistemas Informáticos, departamento de Ingeniería Informática, Universidad de Burgos.		
Web			
Descripción General	Conocer y aplicar los modelos de programación utilizados en el desarrollo de algoritmos y aplicaciones de tipo Big Data.		



1. Situación / Sentido de la asignatura

1.1 Contextualización

La asignatura se engloba dentro de la materia de Tecnologías Informáticas para el Big Data. En esta asignatura se presenta el paradigma de la programación funcional como una de las bases para los modelos de programación de Big Data actuales. En concreto, se explora el modelo map/reduce y una de las tecnologías que lo implementan llamada Apache Spark.

1.2 Relación con otras asignaturas

La asignatura está relacionada principalmente con las de la materia a la que pertenece: Tecnologías Informáticas para el Big Data. En concreto, las asignaturas más relacionadas son:

- Infraestructuras para el Big Data.
- Arquitecturas para el Big Data.

La relación se debe a que en estas asignaturas se presentan las infraestructuras y las arquitecturas que hacen posible los modelos de programación para el Big Data.

1.3 Prerrequisitos

Nivel de inglés medio, dado que la mayoría de los materiales complementarios estarán en dicho idioma.

Se necesitan conocimientos de programación, puesto que se presentarán los modelos de programación para el Big Data y el alumno deberá realizar prácticas y trabajos sobre algún lenguaje de programación de los que soporta Spark.



2. Competencias

2.1 Generales del título

CG1. Adquisición de competencias teóricas y prácticas para el análisis y diseño de soluciones empresariales en Big Data (almacenamiento y procesamiento de grandes volúmenes de información heterogénea).

2.2 Específicas materia

CBD1. Capacidad de diseñar e implementar sistemas de descubrimiento de conocimiento en grandes bases de datos distribuidas.



3. Resultados de aprendizaje

Al finalizar la asignatura, el alumno será capaz de:

- Conocer las restricciones y exigencias a las que están sometidos los programas que trabajan con Big Data.
- Programar de acuerdo al paradigma de la programación funcional.
- Desarrollar algoritmos sobre el modelo de programación Map/Reduce.
- Implementar soluciones sobre plataformas basadas en Map/Reduce, como por ejemplo Spark.



4. Contenido / Programa de la asignatura

4.1 Unidades docentes (bloques de contenidos)

Tema 1: Introducción a Big Data, Map/Reduce y Spark

- Introducción a los modelos de programación para Big Data. Se presenta qué es Big Data, qué limitaciones y requerimientos tiene y se analizan algunas de las tecnologías de Big Data más populares actualmente.
- Introducción a la programación funcional. Se hace un repaso al paradigma de la programación funcional. Se analizan ejemplos de cómo funciona y se introduce el lenguaje de programación Scala.
- Introducción a Map/Reduce. Etapa map y etapa reduce. Los modelos de programación para Big Data actuales se centran en el modelo MapReduce en el cual todo el procesamiento se realiza mediante sucesivas etapas de mapeo y reducción de los datos. Se ilustra el modelo con ejemplos prácticos y sus aplicaciones.
- Introducción a Spark. Se presenta Apache Spark, el framework de procesamiento de datos en Big Data más utilizado actualmente. Se presentan sus motivaciones, ventajas frente a otros frameworks y por qué goza de su reputación.

Tema 2: Resilient Distributed Dataset

- Estructura y principal abstracción de Spark: RDD y las acciones/transformaciones a realizar sobre ellas. La principal abstracción de Spark son los Resilient Distributed Datasets que son la estructura de datos sobre la que reposa todo el funcionamiento de Spark. Estas estructuras distribuidas permiten realizar acciones y transformaciones sobre ellas que serán analizadas y explicadas en detalle.

Tema 3: Particionado y barajado

- Se presenta cómo Spark distribuye los datos por el clúster de cómputo y cómo esto afecta a su rendimiento.

Tema 4: Spark SQL

- Se presenta Spark SQL que es el módulo de Spark para el trabajo sobre datos estructurados, en concreto se presentan los Datasets y Dataframes que se construyen sobre la abstracción de los RDDs. Se aprende a trabajar con ellos así como las transformaciones y acciones a realizar sobre ellos.

4.2 Bibliografía

- H. Karau, A. Konwinski, P. Wendell, and M. Zaharia. Learning Spark: lightning-fast big data analysis. O' Reilly Media, Inc., 2015



5. Metodología de enseñanza y dedicación del estudiante a la asignatura

Actividad Formativa	Competencias relacionadas	Horas	Presencialidad (%)
Clases, conferencias y técnicas expositivas		12	0
Actividades autónomas y en grupo (trabajos y lecturas dirigidas)		45	0
Pruebas de seguimiento y exposición de trabajos		10	50
Tutoría individual, participación en foros y otros medios colaborativos		8	0



6. Temporalización (por bloques temáticos)

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Introducción al Big Data, MapReduce y Spark	1	18/10 – 24/10
Principales abstracciones de Spark	1	25/10 – 7/11
Spark SQL	1	8/11 – 21/11



7. Evaluación

Instrumento / Procedimiento	Peso primera convocatoria	Peso segunda convocatoria
Evaluación sumativa, que incluye pruebas parciales individuales y prueba final	30%	30%
Realización de proyectos, resolución de problemas y casos	40%	40%
Realización de trabajos	20%	20%
Participación en foros y otros medios participativos	10%	10%

Crterios / Comentarios a la evaluación

- **Convocatoria ordinaria:** La calificación final será la media ponderada al porcentaje indicado en las tablas. Para la superación de la asignatura se exigirá un mínimo de 4 puntos sobre 10 en los procedimientos: "Evaluación sumativa...", "Realización de proyectos..." y "Realización de trabajos".
- **Convocatoria extraordinaria:** La calificación final será la media ponderada al porcentaje indicado en las tablas. Para la superación de la asignatura se exigirá un mínimo de 4 puntos sobre 10 en los procedimientos: "Evaluación sumativa...", "Realización de proyectos..." y "Realización de trabajos".
- Es posible que el procedimiento "Participación en foros y otros medios participativos" no sea recuperable en su totalidad en 2ª convocatoria. La evaluación se basa en la interacción entre los alumnos y es posible que esta no pueda organizarse de forma satisfactoria por restricciones de tiempo o de número de alumnos en ese período. En ese caso, se conservará la nota obtenida en la 1ª convocatoria.



8. Recursos de aprendizaje y apoyo tutorial del curso *online*

Transparencias.
Vídeos.
Enunciados de ejercicios.
Ejercicios resueltos.
Páginas webs relacionadas.
Bibliografía disponible en la Biblioteca.
Tutorías individualizadas o en grupo a demanda de los alumnos.



9. Consideraciones / Comentarios adicionales
