

**Proyecto docente de la asignatura**

Asignatura	Fundamentos de Big Data		
Materia	Análisis de Datos y Big Data		
Módulo			
Titulación	Master Universitario de Investigación en Tecnologías de la Información y las Comunicaciones		
Plan		Código	
Periodo de impartición	Primer Cuatrimestre	Tipo/Carácter	Optativa
Nivel/Ciclo		Curso	1
Créditos ECTS	5		
Lengua en que se imparte	Castellano		
Profesor/es responsable/s	Evaristo J. Abril Domingo, Juan Pablo de Castro Fernández, Javier Manuel Aguiar Pérez y Juan Carlos Aguado Manzano		
Datos de contacto (E-mail, teléfono...)	ejabril@tel.uva.es		
Horario de tutorías	Los fijados por los profesores		
Departamento	Teoría de la Señal y Comunicaciones e Ingeniería Telemática		



1. Situación / Sentido de la Asignatura

1.1 Contextualización

La **cantidad de datos** que se está generando a nivel mundial está **creciendo exponencialmente**. Esto supone una oportunidad para **traducir esos datos en información útil y extraer un enorme valor añadido** en ámbitos muy diversos. La gran cantidad de aplicaciones del análisis de datos y del Big Data hace que se traduzca en una **demanda creciente de profesionales** en estas áreas.

Esta asignatura pretende poner de manifiesto los fundamentos de estas técnicas para proporcionar al alumno una visión global de este nuevo camp.

1.2 Relación con otras materias

Esta asignatura se complementa con otras del este máster como son:

- Fundamentos de Aprendizaje Automático
- Aprendizaje Automático Avanzado
- Arquitecturas Paralelas y Deep Learning

1.3 Prerrequisitos

Conocimientos de bases de datos relacionales.

Conocimiento de SQL.

2. Competencias

2.1 Generales

- Capacidad crítica hacia el conocimiento actual como medio imprescindible para la detección de nuevos retos a resolver y por eso evaluar crítica y constructivamente resultados de investigación de otros. [CG 1]
- Capacidad de iniciar la inmersión conceptual y práctica en el contexto socio-económico, y especialmente en los aspectos que influyen a la transferencia tecnológica del conocimiento, así como en los procesos típicos del ciclo de innovación tecnológica y empresarial. [CG 2]
- Capacidad de analizar y aplicar los conocimientos técnicos específicos de su área en nuevos entornos y contextos, teniendo en cuenta los parámetros y variables más significativas de cada nueva situación. [CG 5]
- Capacidad de entender las implicaciones éticas y sociales de las decisiones adoptadas durante el ejercicio de las labores profesionales y de investigación. [CG 6]
- Capacidad de comprender los factores que generan problemas relacionados con la igualdad de sexo, raza o religión, así como la cultura de paz, dentro del sistema global de I+D+i, así como poder integrar soluciones a estos problemas en las propuestas técnicas. [CG 7]
- Capacidad de comunicar los resultados de investigación mediante artículos escritos, propios de divulgación del conocimiento en el sistema de investigación regido por el sistema de revisión entre pares, o en otros términos, escribir artículos técnicos correctos tanto en el fondo como en la forma. [CG 8]
- Capacidad de emplear las técnicas y medios más adecuados para la comunicación oral en diversos foros de la comunidad académica, científica o empresarial, así como para su divulgación en general en la sociedad, o en otros términos, preparar y realizar presentaciones orales correctas ante audiencias expertas y en contextos divulgativos. [CG 9]
- Capacidad de conocer y emplear técnicas y herramientas relacionadas con el modelado, simulación, experimentación y validación de las propuestas técnicas, así como evaluarlas mediante unos parámetros de bondad establecidos. [CG 10]
- Capacidad de desarrollar la capacidad de aprendizaje y trabajo en grupo tanto en entornos conocidos y restringidos, así como en consorcios internacionales en los que intervienen factores culturales. [CG 11]
- Capacidad de proseguir en un aprendizaje a lo largo de toda la vida (*Life Long Learning*) a través de la asimilación de las técnicas y actitudes propias del trabajo autónomo y auto-dirigido. [CG 13]
- Capacidad de emplear por lo menos un idioma extranjero, preferentemente el inglés, como medio de comunicación oral y escrita dentro de su participación en la comunidad científico-tecnológica internacional. [CG 14]

2.2 Específicas

- Capacidad para describir los conceptos fundamentales asociados a las áreas de análisis de datos y Big Data. [CE-AD 6]
- Capacidad para describir e identificar aplicaciones, estrategias y casos de uso basados en análisis de datos y Big Data en distintos sectores de actividad, así como para explicar las ventajas, los peligros, los desafíos y las barreras de adopción, incluyendo las implicaciones éticas y legales. [CE-AD 7]
- Capacidad para describir las principales tecnologías y herramientas asociadas a Big Data (incluyendo sistemas NoSQL, y los ecosistemas Hadoop y Spark) y realizar un uso básico de una selección de las mismas. [CE-AD 8]
- Capacidad para comprender las técnicas y teorías clave sobre visualización de datos y utilizar al menos una herramienta de visualización de datos para generar visualizaciones eficientes. [CE-AD 9]

3. Objetivos

Los objetivos de esta asignatura son:

- Describir los conceptos fundamentales asociados al área de Big Data
- Describir e identificar aplicaciones, estrategias y casos de uso de Big Data y Datos Abiertos.
- Conocer las implicaciones legales y éticas del Big Data
- Capacidad para describir distintos tipos de sistemas NoSQL y sus características.
- Capacidad para comparar las características de distintos sistemas de bases de datos NoSQL entre sí y con los sistemas relacionales.
- Capacidad para identificar en qué situaciones utilizar un sistema de bases de datos relacional y en cuales una solución NoSQL.
- Capacidad para realizar operaciones de consulta de datos en sistemas NoSQL.
- Capacidad para diseñar, implementar y operar sistemas de bases de datos NoSQL.
- Capacidad para describir los conceptos fundamentales asociados a las áreas de análisis de datos y Big Data.
- Capacidad para describir e identificar aplicaciones, estrategias y casos de uso basados en análisis de datos y Big Data en distintos sectores de actividad, así como para explicar las ventajas, los peligros, los desafíos y las barreras de adopción, incluyendo las implicaciones éticas y legales.
- Capacidad para describir las principales tecnologías y herramientas asociadas a Big Data (incluyendo sistemas NoSQL, y los ecosistemas Hadoop y Spark) y realizar un uso básico de una selección de las mismas.
- Capacidad para desarrollar una visualización efectiva, según los estándares más aceptados.
- Capacidad de elegir la técnica o técnicas de visualización más adecuadas para el análisis de un conjunto de datos particular o para la presentación de resultados de un análisis.
- Capacidad para utilizar como herramienta de visualización Tableau.

4. Contenidos

Los contenidos de la asignatura Fundamentos de Big Data son:

Bloque 0:

- Introducción a Big Data
- Aplicaciones de Big Data
- Aspectos legales y éticos
- Datos abiertos

Bloque 1 y 2:

- Fundamentos de tecnologías Big Data (SQL, NoSQL, ecosistemas Hadoop y Spark)

Bloque 3:

- Visualización de datos

Bloque 0: Introducción

Carga de trabajo en créditos ECTS: 0.4

a. Contextualización y justificación

Un analista o científico de datos debe ser capaz de utilizar diferentes formas de representar los datos tanto para realizar el análisis de los mismos como para transmitir los resultados más importantes que se encuentran después de un análisis.

b. Objetivos de aprendizaje

Los objetivos de este bloque 0 son:

- Describir los conceptos fundamentales asociados al área de Big Data
- Describir e identificar aplicaciones, estrategias y casos de uso de Big Data y Datos Abiertos.
- Conocer las implicaciones legales y éticas del Big Data

c. Contenidos

- Introducción a los Fundamentos de Big Data:
 - Introducción
 - Algunos Conceptos Importantes
 - ¿Qué es Big Data?
 - Campos de Aplicación de la Ciencia de Datos y Big Data
- Aspectos Legales y Éticos del Big Data
 - Introducción
 - Aspectos Legales del Big Data
 - Aspectos Éticos del Big Data
 - Datos Abiertos

d. Métodos docentes

Clase magistrales.

e. Evaluación

Entrega de un trabajo relacionados con uno de los temas del Bloque 0.



f. Bibliografía básica

Material proporcionado por el profesor.

g. Bibliografía complementaria

h. Recursos necesarios



Bloque 1: Almacenamiento y consulta de datos

Carga de trabajo en créditos ECTS: 1,8

a. Contextualización y justificación

Un analista o científico de datos debe conocer cómo almacenar los datos y cómo acceder a los mismos para poder procesarlos y obtener valor de los mismos. Esta asignatura tiene por tanto como objetivo analizar las arquitecturas de almacenamiento de datos y muy especialmente los sistemas de bases de datos, tanto convencionales como no convencionales. Así pues, los fundamentos de los sistemas de bases de datos no convencionales NoSQL (como MongoDB o Cassandra) son objeto de esta asignatura.

b. Objetivos de aprendizaje

AC.1 Capacidad para describir distintos tipos de sistemas NoSQL y sus características.

AC.2 Capacidad para comparar las características de distintos sistemas de bases de datos NoSQL entre sí y con los sistemas relacionales.

AC.3 Capacidad para identificar en qué situaciones utilizar un sistema de bases de datos relacional y en cuales una solución NoSQL.

AC.4 Capacidad para realizar operaciones de consulta de datos en sistemas NoSQL.

AC.5 Capacidad para diseñar, implementar y operar sistemas de bases de datos NoSQL.

c. Contenidos

1. Introducción a los Sistemas NoSQL
 - Motivación
 - Comparación con las Bases de datos relacionales
 - Teorema CAP
 - Tipos: Sistemas clave-valor, BBDD orientadas a columnas, BBDD orientadas a documentos, BBDD orientadas a grafos
2. BBDD orientadas a columnas
 - Generalidades
 - Ejemplos: Cassandra
3. BBDD orientadas a Clave-Valor
 - DynamoDB
4. BBDD orientadas a documentos
 - Generalidades
 - Ejemplo: MongoDB
5. BBDD orientadas a grafo
 - Neo4j

d. Métodos docentes



Clase magistral.

Seminarios de debate sobre los encargos.

e. Evaluación

4 prácticas relacionadas con cada base de datos.

La calificación será la media aritmética de la recibida en cada entrega.

f. Bibliografía básica

Material proporcionado por el profesor.

Documentación de los sitios web de cada base de datos.

g. Bibliografía complementaria

h. Recursos necesarios



Bloque 2: Fundamentos de tecnologías Big Data

Carga de trabajo en créditos ECTS: 1,8

a. Contextualización y justificación

Un analista o científico de datos debe conocer las distintas tecnologías Big Data que permiten el almacenamiento y procesamiento de grandes volúmenes de datos a la velocidad adecuada y en el instante oportuno. Por tanto, el objetivo de este Bloque consiste en el estudio y análisis de las diferentes tecnologías emergentes que conforman el ecosistema Big Data (Hadoop, MapReduce, etc.).

b. Objetivos de aprendizaje

- Capacidad para describir los conceptos fundamentales asociados a las áreas de análisis de datos y Big Data.
- Capacidad para describir e identificar aplicaciones, estrategias y casos de uso basados en análisis de datos y Big Data en distintos sectores de actividad, así como para explicar las ventajas, los peligros, los desafíos y las barreras de adopción, incluyendo las implicaciones éticas y legales.
- Capacidad para describir las principales tecnologías y herramientas asociadas a Big Data (incluyendo sistemas NoSQL, y los ecosistemas Hadoop y Spark) y realizar un uso básico de una selección de las mismas.

c. Contenidos

- Fundamentos de Big Data
 - Definiciones y conceptos
 - Características
 - Beneficios e inconvenientes
 - Aplicaciones
 - Tipos de datos
 - Habilidades y perfiles del científico de datos
 - Evolución
 - Ecosistema
- Infraestructura de Big Data
 - Hadoop
 - Características
 - Necesidad
 - Componentes
 - Arquitectura
 - Modos
 - Evolución
 - Ecosistema
 - Distribuciones

d. Métodos docentes

Clase magistral.

Seminarios de seguimiento de las prácticas propuestas.

e. Evaluación

4 prácticas relacionadas con las tecnologías Big Data.

La calificación será la media aritmética de la recibida en cada entrega.



f. Bibliografía básica

Material proporcionado por el profesor.

Documentación de los sitios web de las diferentes tecnologías Big Data.

g. Bibliografía complementaria

h. Recursos necesarios



Bloque 3: Visualización de datosCarga de trabajo en créditos ECTS:

1

a. Contextualización y justificación

Un analista o científico de datos debe ser capaz de utilizar diferentes formas de representar los datos tanto para realizar el análisis de los mismos como para transmitir los resultados más importantes que se encuentran después de un análisis.

b. Objetivos de aprendizaje

- Capacidad para desarrollar una visualización efectiva, según los estándares más aceptados.
- Capacidad de elegir la técnica o técnicas de visualización más adecuadas para el análisis de un conjunto de datos particular o para la presentación de resultados de un análisis.
- Capacidad para utilizar como herramienta de visualización Tableau.

c. Contenidos

- Introducción a la visualización. Concepto
- Fases de creación de una visualización
- Manejo de la herramienta Tableau para realizar visualización.

d. Métodos docentes

Clase magistrales, prácticas y de laboratorio.

e. Evaluación

Entrega de varios problemas de uso de la herramienta de visualización.

f. Bibliografía básica

Material proporcionado por el profesor.

g. Bibliografía complementaria

- "Data Visualization. A handbook for Data Driven Design", Andy Kirk, SAGE Publications, 2016
- "The Visual Display of Quantitative Information", Edward R. Tufte, Graphic Press LLC, Second Edition, 2015
- "Storytelling with data. A data visualization guide for business professionals" Cole Nussbaumer Knaflic, Wiley, 2015

h. Recursos necesarios

Licencia de estudiante de Tableau



5. Métodos docentes y principios metodológicos

Ver apartado 4 del presente Proyecto Docente

6. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teóricas	30	Estudio y trabajo autónomo individual	50
Clases prácticas		Estudio y trabajo autónomo grupal	25
Laboratorios	20		
Prácticas externas, clínicas o de campo			
Seminarios			
Otras actividades			
Total presencial	50	Total no presencial	75



7. Sistema y características de la evaluación

Ver apartado 4 del presente Proyecto Docente

8. Consideraciones finales

