



## Máster Oficial en Ingeniería Informática

### DATOS DE LA ASIGNATURA

**Nombre:**

Big Data

**Denominación en inglés:**

Big Data

**Código:**

1140217

**Carácter:**

Optativo

**Horas:****Totales****Presenciales****No presenciales****Trabajo estimado:**

50

20

30

**Créditos:****Grupos reducidos****Grupos grandes****Aula estándar****Laboratorio****Prácticas de campo****Aula de informática**

1

0

0

0

1

**Departamentos:****Áreas de Conocimiento:**

Tecnologías de la Información

Ciencia de la Computación e Inteligencia Artificial

**Curso:****Cuatrimestre:**

2º - Segundo

Primer cuatrimestre

### DATOS DE LOS PROFESORES

**Nombre:****E-Mail:****Teléfono:****Despacho:**

\*Peregrín Rubio, Antonio

peregrin@uhu.es

87653

TUPB-23

\*Profesor coordinador de la asignatura

**1. Descripción de contenidos****1.1. Breve descripción (en castellano):**

Big Data trata sobre las tecnologías empleadas en esta actual óptica del tratamiento y gestión de grandes volúmenes de datos provenientes de fuentes heterogéneas como son las redes sociales, el correo electrónico, los archivos de video, etc., en volumen ilimitado, estructurados, sin estructura o estados intermedios, para proporcionar a las organizaciones análisis que permitan adelantarse a los acontecimientos tales como las tendencias de mercado. En este sentido, se diferencia del Business Intelligence, orientado al análisis de datos para mostrar cómo funciona un negocio en sus distintas áreas y facilitar decisiones, usando normalmente datos estructurados provenientes básicamente del Data Warehouse de la empresa. En esta asignatura se definirá qué es esta tecnología emergente, qué herramientas y métodos se utilizan para implementarla en las organizaciones, sus aplicaciones, casos reales y modelos de implantación, etc.

Sus contenidos por tanto abarcarían:

- Por qué surge el Big Data: Mundo de datos.
- Definiciones de Big Data, impacto en la sociedad actual, etc. Necesidades de nuevas arquitecturas, técnicas, algoritmos y análisis. Volumen, velocidad y variedad de datos que motiva el Big Data.
- Herramientas y métodos en Big Data: Plataformas (Ej.: Hadoop (Open access)), Librerías (Ej.: Mahout, Nimble, SystemML, Ricardo, Rhipe), Paradigmas de programación (Ej.: MapReduce), Algoritmos, etc.
- Aplicaciones técnicas e implantaciones reales. Ejemplos típicos: telefonía, genómica, astronomía, transacciones bancarias, tráfico en internet, información web, etc.
- Business Intelligence y Data Analytics frente a Big Data: diferencias entre estos términos, a veces confundidos y frecuentemente relacionados.
- Aspectos y consideraciones éticas (privacidad, etc).
- El futuro del Big Data: Áreas en evolución, tareas pendientes, hibridación con otras tecnologías, etc.

**1.2. Breve descripción (en inglés):**

Big Data course deals with the technologies used in the current point of view of data science: treatment and management of large volumes of data from heterogeneous sources like social networks, email, video files, etc., in unlimited volume, structured and unstructured, in order to provide the analysis that let organizations to anticipate events such as market trends, etc. In this sense, it differs from Business Intelligence, oriented to data analysis, to show how a business works in different areas and facilitate decisions, usually using data basically from the Data Warehouse structured company. In this course, it will be defined what is this emerging technology, which tools and methods are used to implement in organizations, applications, case studies and deployment models, etc.

Therefore its contents encompass:

- World of data: Why Big Data arises.
- Big Data definitions, impact on today's society, etc. Needs new architectures, techniques, algorithms and analysis. Volume, velocity and variety of data that motivates the Big Data.
- Tools and methods in Big Data: Platforms (Ex.: Hadoop (Open Access)), Book (eg. Mahout, Nimble, SystemML, Ricardo, Rhipe) Programming paradigms (eg.: MapReduce), algorithms, etc.
- Technical applications and actual implementations. Typical examples: telephone, genomics, astronomy, banking, Internet traffic, website information, etc. Business Intelligence and Data Analytics vs. Big Data: differences between these terms, sometimes confused and often related.
- Aspects and ethical considerations (privacy, etc.).
- The Future of Big Data: Areas evolving, hybridization with other technologies, etc.

**2. Situación de la asignatura****2.1. Contexto dentro de la titulación:**

Se trata de una asignatura que se orienta a la demanda de especialización en Ciencia de los Datos, situada en el segundo cuatrimestre del segundo curso de Máster, y de forma optativa.

**2.2. Recomendaciones:**

Ninguna.

**3. Objetivos (Expresados como resultados del aprendizaje):**

Los objetivos de esta asignatura son:

- Que el alumno sepa qué es la ciencia de los datos y el big data, su relevancia hoy en día. Conocerá además la relación que tiene éste ámbito con otros tales como Internet of Things, Smart Cities, Minería de Datos, Business Intelligence, etc.
- Que tenga la capacidad de comprender un problema de Big Data y aplicar los principios y conocimientos de computación para afrontar su resolución
- Que además, sepa en qué recursos se basa resolver problemas de este ámbito, es decir, las del ámbito estadístico y especialmente las del ámbito de la Inteligencia Computacional.
- Que conozca en realidad qué técnicas están desarrolladas para resolver problemas de Big Data, es decir, más allá de la teoría de la necesidad y la potencialidad del Big Data, qué puede utilizar y para qué.

#### 4. Competencias a adquirir por los estudiantes

##### 4.1. Competencias específicas:

##### 4.2. Competencias básicas, generales o transversales:

- **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
- **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
- **CG1:** Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la ingeniería informática
- **CG8:** Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos

## 5. Actividades Formativas y Metodologías Docentes

### 5.1. Actividades formativas:

- Sesiones de Teoría sobre los contenidos del Programa.
- Sesiones de Resolución de Problemas.
- Sesiones Prácticas en Laboratorios Especializados o en Aulas de Informática.
- Actividades Académicamente Dirigidas por el Profesorado: seminarios, conferencias, desarrollo de trabajos, debates, tutorías colectivas, actividades de evaluación y autoevaluación.

### 5.2. Metodologías docentes:

- Clase Magistral Participativa.
- Desarrollo de Prácticas en Laboratorios Especializados o Aulas de Informática en grupos reducidos.
- Resolución de Problemas y Ejercicios Prácticos.
- Tutorías Individuales o Colectivas. Interacción directa profesorado-estudiantes.
- Planteamiento, Realización, Tutorización y Presentación de Trabajos.
- Conferencias y Seminarios.
- Evaluaciones y Exámenes.

### 5.3. Desarrollo y justificación:

- Sesiones de teoría, donde exponer los principios teóricos: el profesor, siguiendo el programa de la asignatura, explica, mediante presentaciones, pizarra y de viva voz, los contenidos teóricos de la asignatura.
- Sesiones de laboratorio. En aulas de informática, para resolver problemas prácticos que se proponen a los alumnos... se les da un tiempo para resolverlos, y se puntúan los resultados de los trabajos recogidos en tiempo y forma.
- Trabajo en grupos reducidos. Problemas planteados y prácticas que los alumnos no resuelvan de forma individual sino en grupo.
- Resolución y entrega de problemas/prácticas. El profesor podría requerir al alumno entregar en determinados momentos, prácticas o problemas para evaluarlos dentro de los criterios y niveles enunciados en la parte correspondiente a criterios de evaluación de esta guía de la asignatura.

## 6. Temario desarrollado:

- Tema 1: Introducción a la Ciencia de los Datos y el Big Data  
Repaso a la terminología: Minería de Datos, Bussines Intelligence, Internet of Things, Cloud Computing, Data warehouse, Big Analytics,... etc, y la relación de cada uno de ellos con Big Data). ¿Por qué tiene tanto auge fuera del campo de la propia computación el Big Data?. Ejemplos y Aplicaciones.
- Tema 2: Introducción al Paradigma de Computación MapReduce  
MapReduce en detalle. Comparación con otros modelos de computación paralela.
- Tema 3: Apache Hadoop.  
Apache Hadoop: Versión 1, Versión 2, Ecosistema.
- Tema 4: Recursos de Inteligencia Computacional, Machine Learning y Data Mining en Big Data  
Estudio de los recursos y retos en éste área. Biblioteca Mahout.
- Tema 5: Introducción a Spark.  
Estudio de Spark, fundamentos, ecosistema, etc.
- Tema 6: Nuevos Retos en Big Data

## 7. Bibliografía

### 7.1. Bibliografía básica:

- Big data : a revolution that will transform how we live, work, and think, Viktor Mayer-Schönberger, Kenneth Cukier, HoughtonMifflin Harcourt, 2013 (Versión Castellano: Big Data, La revolución de los Datos Masivos)
- Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. IBM. Paul Zikopoulos, Chris Eaton. McGraw Hill Professional, 29/05/2015. <http://www-01.ibm.com/software/data/bigdata/>
- Analítica predictiva: Predecir el futuro utilizando Big Data / Eric Siegel. Madrid : Anaya Multimedia, 2013

### 7.2. Bibliografía complementaria:

- DEL CLOUD COMPUTING AL BIG DATA: Vision introductoria para jovenes emprendedores. Jordi Torres i Vinnals. Editorial UOC – PID\_00194204. Primera edicion: septiembre 2012. Todos los derechos reservados de esta edicio?n, FUOC, 2012. v.3.0 Espan?a de Creative Commons.
- Hadoop, Soluciones Big Data,, Boris Lublinsky, Kevin Smith, Alex Yakubovich, Anaya Multimedia 2014
- Big Data, Técnicas, Herramientas y Aplicaciones, María Pérez Marqués, 2015

## 8. Sistemas y criterios de evaluación.

### 8.1. Sistemas de evaluación:

- Examen de teoría/problemas
- Defensa de Prácticas
- Defensa de Trabajos e Informes Escritos
- Seguimiento Individual del Estudiante
- Examen de prácticas

### 8.2. Criterios de evaluación y calificación:

La nota se calculará puntuando estos cinco ítems, que acumulan la ponderación del Sistema de Evaluación de la ficha:

- Examen de Teoría/Problemas: 25%
- Defensa de Prácticas: 30%
- Examen de Prácticas: 20%
- Defensa de Trabajos e Informes Escritos: 20%
- Seguimiento Individual del Estudiante: 5%

Los tres ítems son independientes (no necesitan nota mínima), y su nota se mantiene hasta Septiembre del curso inclusive. Se garantiza la adquisición de las competencias de la siguiente forma: mediante el examen de teoría/problemas, las competencias: CB7 y CG8 ; mediante las páticas de laboratorio, las competencias: CB7, CG1 y CG8; mediante el examen de prácticas, las competencias: CB7; mediante la defensa de trabajos e informes escritos, las competencias: CB7 y CB9; y mediante el seguimiento individual delestudiante, la competencia: CB9.

