

**ANX-PR/CL/001-01**  
**GUÍA DE APRENDIZAJE**

**ASIGNATURA**

Big data

**CURSO ACADÉMICO - SEMESTRE**

2016-17 - Primer semestre

## Datos Descriptivos

<b>Nombre de la Asignatura</b>	Big data
<b>Titulación</b>	10AN - Master Universitario en Ingeniería Informática
<b>Centro responsable de la titulación</b>	Escuela Técnica Superior de Ingenieros Informáticos
<b>Semestre/s de impartición</b>	Tercer semestre
<b>Módulos</b>	Intensificación en informática Intensificación en informática
<b>Materias</b>	Computación científica Sistemas y servicios distribuidos
<b>Carácter</b>	Optativa
<b>Código UPM</b>	103000640
<b>Nombre en inglés</b>	Big data

## Datos Generales

<b>Créditos</b>	6	<b>Curso</b>	2
<b>Curso Académico</b>	2016-17	<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Inglés	<b>Otros idiomas de impartición</b>	

## Requisitos Previos Obligatorios

### Asignaturas Previas Requeridas

El plan de estudios Master Universitario en Ingeniería Informática no tiene definidas asignaturas previas superadas para esta asignatura.

### Otros Requisitos

El plan de estudios Master Universitario en Ingeniería Informática no tiene definidos otros requisitos para esta asignatura.

## Conocimientos Previos

### Asignaturas Previas Recomendadas

Sistemas cognitivos

Sistemas interactivos

Computación para ciencias e ingeniería

### Otros Conocimientos Previos Recomendados

El coordinador de la asignatura no ha definido otros conocimientos previos recomendados.



CAMPUS  
DE EXCELENCIA  
INTERNACIONAL

**UNIVERSIDAD POLITÉCNICA DE MADRID**  
Escuela Técnica Superior de Ingenieros Informáticos  
**PROCESO DE COORDINACIÓN DE LAS ENSEÑANZAS**

ANX-PR/CL/001-01: GUÍA DE APRENDIZAJE



Código PR/CL/001

## Competencias

---

- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
- 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
- CE1 - Capacidad para la integración de tecnologías, aplicaciones, servicios y sistemas propios de la Ingeniería Informática, con carácter generalista, y en contextos más amplios y multidisciplinares.
- CE10 - Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería.
- CE13 - Capacidad para utilizar y desarrollar metodologías, métodos, técnicas, programas de uso específico, normas y estándares de computación gráfica.
- CE16 - Habilidad para hacer conexiones entre los deseos y necesidades del consumidor o cliente y lo que la tecnología puede ofrecer
- CE4 - Capacidad para modelar, diseñar, definir la arquitectura, implantar, gestionar, operar, administrar y mantener aplicaciones, redes, sistemas, servicios y contenidos informáticos.
- CE9 - Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.
- CG12 - Capacidad de trabajar de forma independiente en su campo profesional
- CG9 - Apreciación de los límites del conocimiento actual y de la aplicación práctica de la tecnología más reciente

## Resultados de Aprendizaje

---

- RA108 - Conocer técnicas de visualización y procesos de análisis de datos, y de programación, diseño y depuración de algoritmos, para computación de altas prestaciones.
- RA39 - Ser capaz de procesar datos masivos
- RA41 - Conocer cómo se aplican las técnicas de computación científica en algún campo específico de ciencia o ingeniería
- RA40 - Conocer técnicas de visualización y proceso de análisis de datos

## Profesorado

---

### Profesorado

Nombre	Despacho	e-mail	Tutorías
Latorre De La Fuente, Antonio <b>(Coordinador/a)</b>	4202	a.latorre@upm.es	
Toharia Rabasco, Pablo	4102	pablo.toharia@upm.es	
Montes Sanchez, Jesus	4204	jesus.montes@upm.es	

**Nota.-** Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

## Descripción de la Asignatura

---

Esta asignatura capacitará al alumno para adquirir los fundamentos necesarios para el análisis y la visualización analítica de grandes volúmenes de datos. Con un enfoque eminentemente práctico se presentarán las tecnologías y fundamentos necesarios para llevar a cabo satisfactoriamente el proceso completo de análisis de datos en el contexto de Big Data, desde los datos en crudo hasta su visualización, pasando por los modelos derivados de los mismos.

## Temario

---

1. Introducción y fundamentos
  - 1.1. Arquitecturas y aplicaciones
  - 1.2. Tipologías de datos
  - 1.3. Visual analytics
2. El ecosistema Big Data
  - 2.1. Introducción al Big Data
  - 2.2. Paradigma Map Reduce
  - 2.3. Hadoop
3. Tecnologías Big Data
  - 3.1. Introducción a los esquemas Big Data típicos
  - 3.2. Introducción a Spark
  - 3.3. Introducción a Tez
  - 3.4. Introducción a Flink
4. Spark
  - 4.1. Estructuras de datos
    - 4.1.1. RDDs (Spark)
    - 4.1.2. Data Frames (SparkSQL)
  - 4.2. Análisis de datos con Spark
    - 4.2.1. MLLib
    - 4.2.2. SparkML

5. Visualización de la información

- 5.1. Fundamentos de la visualización de la información
- 5.2. Abstracciones de datos
- 5.3. Abstracciones de tareas
- 5.4. Técnicas interactivas y codificación visual
- 5.5. Métodos de diseño
- 5.6. Análisis de ejemplos de sistemas
- 5.7. Lecciones aprendidas

## Cronograma

**Horas totales:** 65 horas

**Horas presenciales:** 65 horas (41.7%)

**Peso total de actividades de evaluación continua:**  
100%

**Peso total de actividades de evaluación sólo prueba final:**  
100%

Semana	Actividad Presencial en Aula	Actividad Presencial en Laboratorio	Otra Actividad Presencial	Actividades Evaluación
Semana 1	<p><b>Tema 1</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 2</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
Semana 2	<p><b>Tema 2</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 1</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 3	<p><b>Tema 3</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Práctica 1</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 4	<p><b>Tema 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 1</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 5	<p><b>Tema 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 2</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		<p><b>Entrega de proyectos prácticos</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Actividad no presencial</p>
Semana 6	<p><b>Tema 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 2</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 7	<p><b>Tema 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Práctica 2</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		



Semana 8	<p><b>Tema 5</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Práctica 2</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 9	<p><b>Tema 5</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 5</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			<p><b>Entrega de proyectos prácticos</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Actividad no presencial</p>
Semana 10	<p><b>Tema 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 11	<p><b>Tema 5</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Práctica 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 12	<p><b>Tema 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 13	<p><b>Tema 5</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Práctica 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 14	<p><b>Tema 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 15	<p><b>Tema 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Tema 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Práctica 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
Semana 16		<p><b>Práctica 3</b> Duración: 04:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		<p><b>Entrega de proyectos prácticos</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Actividad no presencial</p>

Semana 17				<b>Examen</b> Duración: 01:00 EX: Técnica del tipo Examen Escrito Evaluación continua y sólo prueba final Actividad presencial
-----------	--	--	--	--

**Nota.-** El cronograma sigue una planificación teórica de la asignatura que puede sufrir modificaciones durante el curso.

**Nota 2.-** Para poder calcular correctamente la dedicación de un alumno, la duración de las actividades que se repiten en el tiempo (por ejemplo, subgrupos de prácticas") únicamente se indican la primera vez que se definen.

## Actividades de Evaluación

Semana	Descripción	Duración	Tipo evaluación	Técnica evaluativa	Presencial	Peso	Nota mínima	Competencias evaluadas
5	Entrega de proyectos prácticos	00:00	Evaluación continua y sólo prueba final	TG: Técnica del tipo Trabajo en Grupo	No	12%	4 / 10	CG9, CB7, CB10, CE9, CE10, CE13, CE16, CG12, CE1, CE4
9	Entrega de proyectos prácticos	00:00	Evaluación continua y sólo prueba final	TG: Técnica del tipo Trabajo en Grupo	No	28%	4 / 10	CG9, CB7, CB10, CE9, CE10, CE13, CE16, CG12, CE1, CE4
16	Entrega de proyectos prácticos	00:00	Evaluación continua y sólo prueba final	TG: Técnica del tipo Trabajo en Grupo	No	40%	4 / 10	CG9, CB7, CB10, CE9, CE10, CE13, CE16, CG12, CE1, CE4
17	Examen	01:00	Evaluación continua y sólo prueba final	EX: Técnica del tipo Examen Escrito	Sí	20%	4 / 10	CG9, CB7, CB10, CE9, CE10, CE13, CE16, CG12, CE1, CE4

## Criterios de Evaluación

Se describen a continuación los criterios de evaluación considerados en la asignatura. El sistema de evaluación continua será el que se aplicará con carácter general a todos los estudiantes que cursen la asignatura. La guía de aprendizaje se centra por tanto en este sistema y detalla sus actividades de evaluación en los apartados "Evaluación sumativa" y "Cronograma de la asignatura". Las actividades de evaluación del "Sistema de evaluación mediante sólo prueba final" no forman parte de esos apartados y se describen exclusivamente en este apartado de "Criterios de Evaluación".

### Elección del sistema de evaluación

En la convocatoria ordinaria, la elección entre el sistema de evaluación continua o el sistema de evaluación mediante sólo prueba final corresponde al estudiante. Quien desee seguir el sistema de evaluación mediante sólo prueba final, deberá **OBLIGATORIAMENTE** comunicarlo **DURANTE LOS 15 PRIMEROS DÍAS NATURALES**, a contar desde el inicio de la actividad docente de la asignatura, mediante escrito dirigido al **Coordinador de la asignatura**, que entregará dentro del plazo establecido y a través del Registro de la Secretaría de Alumnos. En dicho escrito deberá constar:

"D. \_\_\_\_\_ con DNI \_\_\_\_\_ y nº de matrícula \_\_\_\_\_,

SOLICITA:

Ser evaluado en este semestre mediante el sistema de evaluación mediante sólo prueba final establecido por la siguiente asignatura:

- Asignatura \_\_\_\_\_, titulación \_\_\_\_\_, curso \_\_\_\_\_

Firmado:

"

Esta solicitud sólo se considerará a los efectos del semestre en curso. En posteriores semestres deberá necesariamente ser cursada de nuevo. No obstante lo anterior, cuando exista causa sobrevenida y de fuerza mayor que justifique el cambio del

proceso de evaluación, el estudiante que haya optado (por omisión) por el sistema de evaluación continua podrá solicitar al Tribunal de la Asignatura ser admitido en los exámenes y actividades de evaluación que configuran el sistema de evaluación mediante sólo prueba final. El tribunal de la asignatura, una vez analizadas las circunstancias que se hagan constar en la solicitud, dará respuesta al estudiante con la mayor antelación a la celebración del examen final que sea posible.

**La información completa relativa a este sistema de evaluación puede encontrarla en el siguiente enlace:**

<http://www.fi.upm.es/?pagina=1147>

### **Sistema de evaluación**

La asignatura se evaluará mediante 2 tipos de pruebas:

1. **Examen de la asignatura.** Al final de la asignatura se llevará a cabo un examen sobre todo el contenido de la misma.
2. **Proyectos prácticos.** El enunciado de los mismos se presentará en clase, en la propia aula y durante el horario regular de la asignatura, en las fechas especificadas en el calendario de la asignatura. El desarrollo de estos proyectos se llevará a cabo tanto de forma presencial en las clases prácticas en el laboratorio, como no presencial usando los recursos ofrecidos por el Centro de Cálculo para tal fin, y apoyándose en las tutorías para la resolución de cualquier aspecto vinculado con el desarrollo de los mismos. Los plazos de entrega estarán escalonados a lo largo del desarrollo del curso y se publicarán en la página web de la asignatura con suficiente antelación.

La nota final de la asignatura se calculará considerando un peso de 20% para el examen y otro 80% para la nota media de los proyectos prácticos. Para aprobar la asignatura, además de tener una nota final mayor o igual a 5, habrá que tener una nota mínima de 4 en cada una de las dos partes.

### **Evaluación en la convocatoria extraordinaria de julio**

En caso de que el alumno no haya superado el examen, deberá presentarse al examen de esta convocatoria. Si el alumno no ha superado satisfactoriamente los ejercicios prácticos, tendrá un plazo extraordinario para volver a entregarlos hasta el final del día anterior al de la celebración del examen.

### **Plazos de entrega**

Las fechas exactas establecidas como plazo máximo para la entrega de las distintas prácticas serán publicados en la página Web de la asignatura. Los plazos de entrega estarán escalonados a lo largo del curso con el objetivo de facilitar el desarrollo de las prácticas en paralelo con el temario y maximizar así su beneficio. En ningún caso se admitirán prácticas entregadas fuera de plazo.

## Recursos Didácticos

Descripción	Tipo	Observaciones
Libro 1	Bibliografía	Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006.
Libro 2	Bibliografía	Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley, ISBN: 0321321367, 2005
Libro 3	Bibliografía	Ian Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann, ISBN: 0120884070, 2005.
Libro 4	Bibliografía	Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, ISBN: 978-0-12-374856-0, 2011.
Libro 5	Bibliografía	Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. Mastering the information age. Solving problems with visual analytics 2010 Eurographics Association.
Libro 6	Bibliografía	Tamara Munzner. Visualization Analysis and Design. A K Peters Visualization Series. CRC Press. Nov. 2014.
Libro 7	Bibliografía	Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia. Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly Media. 2015.
Libro 8	Bibliografía	Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills. Advanced Analytics with Spark: Patterns for Learning from Data at Scale. O'Reilly Media. 2015.
Documentación de Spark	Recursos web	<a href="http://spark.apache.org/docs/latest/">http://spark.apache.org/docs/latest/</a>
Aula de clase asignada por Jefatura de Estudios.	Equipamiento	
Sistema del Centro de Cálculo asignado para el desarrollo de los ejercicios prácticos	Equipamiento	
Puestos de trabajo en sala, asignados por Centro de Cálculo.	Equipamiento	
Página web de la asignatura	Recursos web	<a href="http://laurel.datsi.fi.upm.es/docencia/asignaturas/bd">http://laurel.datsi.fi.upm.es/docencia/asignaturas/bd</a>
Documentación de Hive	Recursos web	<a href="https://cwiki.apache.org/confluence/display/Hive/Home">https://cwiki.apache.org/confluence/display/Hive/Home</a>

## Otra Información

Esta asignatura se va a impartir conjuntamente con el alumnado procedente de la iniciativa EIT-Digital, de ahí que a partir del curso 2015/2016 pase a impartirse en inglés.