



## Guía docente de la asignatura

<b>Asignatura</b>	INFRAESTRUCTURA PARA EL BIG DATA	
<b>Materia</b>	TECNOLOGÍAS INFORMÁTICAS PARA EL BIG DATA	
<b>Módulo</b>	(vacío)	
<b>Titulación</b>	MASTER UNIVERSITARIO EN INTELIGENCIA DE NEGOCIO Y BIG DATA EN ENTORNOS SEGUROS	
<b>Plan</b>		<b>Código</b>
<b>Periodo de impartición</b>	PRIMER CUATRIMESTRE	<b>Tipo/Carácter Obligatoria</b>
<b>Nivel/Ciclo</b>	MASTER	<b>Curso 2018-2019</b>
<b>Créditos ECTS</b>	3 ECTS	
<b>Lengua en que se imparte</b>	CASTELLANO	
<b>Profesor/es responsable/s</b>	BENJAMÍN SAHELICES FERNÁNDEZ (responsable teoría y prácticas)	
<b>Datos de contacto (E-mail, teléfono...)</b>	E-MAIL: <a href="mailto:benja@infor.uva.es">benja@infor.uva.es</a>	
<b>Horario de tutorías</b>	Véase <a href="http://www.uva.es">www.uva.es</a> → Grados → Grado en Ingeniería Informática → Tutorías	
<b>Departamento</b>	DEPARTAMENTO DE INFORMÁTICA	

### 1. Situación / Sentido de la Asignatura

#### 1.1 Contextualización

Dentro de la perspectiva de los estudios de un máster sobre informática de perfil profesionalizante, en esta asignatura se pretende, en su primer bloque temático, aportar conocimientos sobre conceptos de programación para la obtención de altas prestaciones en procesamiento de Big Data. En el segundo bloque temático el objetivo es que el estudiante adquiera competencias sobre nuevas tecnologías relacionadas con sistemas distribuidos, cluster escalables de computación, virtualización en clústers y centros de datos y arquitecturas para plataformas cloud.

#### 1.2 Relación con otras materias

La materia guarda relación con asignaturas de programación, sistemas operativos, administración de sistemas, seguridad, evaluación, rendimiento y arquitectura de computadores.

#### 1.3 Prerrequisitos

Conocimientos básicos sobre programación, sistemas operativos y arquitectura de computadores.



## 2. Competencias

### 2.1 Generales del título

Código	Descripción
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

Código	Descripción
CBD1	Capacidad de diseñar e implementar sistemas de descubrimiento de conocimiento en grandes bases de datos distribuidas
CBD2	Capacidad de analizar, diseñar y construir o configurar sistemas de almacenamiento escalable y procesamiento escalable

## 3. Resultados de aprendizaje

- Comprender los principios de la programación de altas prestaciones, en plataformas de computo no convencionales o con nuevos modelos de programación que explotan las capacidades de paralelismo de los sistemas actuales y futuros para el procesamiento de Big Data.
- Conocer las tecnologías emergentes en el ámbito del clúster de computación, la virtualización y los centros de datos.
- Comprender los principios de funcionamiento del clúster implantados sobre centros de datos virtualizados.
- Instalar y administrar una plataforma de virtualización para la computación Big Data.
- Desarrollar software y desplegar aplicaciones en un entorno Cloud.
- Utilizar los diferentes módulos y servicios de un entorno actual Cloud para el desarrollo, despliegue y mantenimiento de aplicaciones escalables.

### Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases, conferencias y técnicas expositivas	12	Actividades autónomas y en grupo (trabajos y lecturas dirigidas)	45
Pruebas de seguimiento y exposición de trabajos	10		
Tutoría individual, participación en foros y otros medios colaborativos	8		
<b>Total presencial</b>	<b>30</b>	<b>Total no presencial</b>	<b>45</b>



## 4. Contenido / Programa de la asignatura

### 4.1 Unidades docentes (bloques de contenidos)

#### Bloque 1: Infraestructura Cloud para el Big Data

Carga de trabajo en créditos ECTS: 1,5

#### Contenidos

##### TEMA 1. Distributed Systems, Cluster, Virtualization and Clouds

###### 1.1 Clusters

- Clusters, grids, peer-to-peer, cloud
- Computer clusters. Design principles.
- MPP clusters
- Job scheduling
- Top500
- Cluster laboratory: Mosix

###### 1.2 Virtualization

- Virtual machines
- Abstraction levels of virtualization
- Hardware virtualization
- OS-level virtualization
- Hypervisor architecture
- Host-based architecture
- Para-virtualization
- Hardware support for virtualization
- Physical vs. virtual clusters
- Fast deployment, migration and scheduling
- Server consolidation in data centers
- Virtual storage management

###### 1.3 Clouds para el Big Data

- Data center virtualization for cloud computing
- Public, private and hybrid clouds
- Network structure of data centers
- Cloud ecosystem
- IaaS
- PaaS
- SaaS
- Cloud security
- Cloud services: web services, SOA and REST, the cloud stack
- Cloud laboratory: OpenStack
- Cloud laboratory: IBM Cloud
- Big Data Laboratory: Hadoop YARN



## Bloque 2: Modelos de computación distribuida orientados al Big Data

---

Carga de trabajo en créditos ECTS:1,5

### Contenidos

---

#### TEMA 2: MapReduce y Hadoop

- 2.1 Modelo de programación MapReduce
- 2.2 Casos prácticos con Hadoop

### 4.2 Bibliografía

---

#### Bibliografía básica

---

- “Distributed and Cloud Computing. From Parallel Processing to the Internet of Things”, Kai Hwang, Geoffrey C. Fox, Jack J. Dongarra, Morgan Kaufmann, 2012. ISBN: 978-0-12-385880-1
- “Guide to Cloud. Principles and Practice”, Richard Hill, Laurie Hirsch, Peter Lake, Siavash Moshiri, Springer, 2013. ISBN: 978-1-4471-4602-5. DOI 10.1007/978-1-4471-4603-2
- “Apache Hadoop YARN. Moving beyond MapReduce and Batch Processing with Apache Hadoop 2”, Arun C. Murthy, Vinod K. Vavilapalli, Addison Wesley Data and Analytics Series, 2014. ISBN 978-0-321-93450-5
- "The Datacenter as a Computer. An Introduction to the Design of Warehouse-Scale Machines. Second Edition.", Luiz André Barroso, Jimmy Clidaras, Urs Holzle, Synthesis Lectures on Computer Architecture, Morgan&Claypool Publishers, 2013. ISBN print: 978-1-60845-009-8. ISBN ebook: 978-1-60845-010-4
- "Virtual Machines. Versatile Platforms for Systems and Processes", James E. Smith and Ravi Nair, Morgan Kaufmann, 2005, ISBN-13: 978-1-55860-910-5. ISBN-10: 1-55860-910-5



**5. Metodología de enseñanza y aprendizaje y su relación con las competencias que debe adquirir el estudiante**

Metodología	Competencia relacionada	Horas de trabajo	Presencialidad (%)
Clases, conferencias y técnicas expositivas	CG1, CBD1, CBD2	12	0
Actividades autónomas y en grupo (trabajos y lecturas dirigidas)	CG1, CBD1, CBD2	45	0
Pruebas de seguimiento y exposición de trabajos	CG1, CBD1, CBD2	10	50
Tutoría individual, participación en foros y otros medios colaborativos	CG1, CBD1, CBD2	8	0

**6. Temporalización (por bloques temáticos)**

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Bloque 1	1,5	Octubre 2018
Bloque 2	1,5	Noviembre 2018



## 7. Sistemas de evaluación

<b>Procedimiento</b>	<b>Peso primera convocatoria</b>	<b>Peso segunda convocatoria</b>
Evaluación sumativa, que incluye pruebas parciales individuales y prueba final	20%	20%
Realización de trabajos, proyectos, resolución de problemas y casos	60%	60%
Participación en foros y otros medios participativos	20%	20%