



Proyecto docente

Asignatura	INFRAESTRUCTURA 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	54545
Periodo de impartición	Primer cuatrimestre	Tipo/Carácter	Obligatoria
Nivel/Ciclo	Máster	Curso	1
Créditos ECTS	3		
Lengua en que se imparte	Castellano. Materiales de referencia en Inglés		
Profesor/es responsable/s	Francisco José Andújar Muñoz		
Datos de contacto (e-mail, teléfono...)	E-MAIL: fandujarm@infor.uva.es		
Horario de tutorías	Véase www.uva.es → Grados → Grado en Ingeniería Informática → Tutorías		
Coordinador	Francisco José Andújar Muñoz		
Departamento	Departamento de informática		
Web			
Descripción General	Conocer las principales infraestructuras hardware y software para la computación Big Data		



1. Situación / Sentido de la asignatura

1.1 Contextualización

Dentro de la perspectiva de los estudios de un máster sobre informática, en esta asignatura se pretende, en su primer bloque temático, aportar conocimientos sobre con sistemas distribuidos, clusters escalables de computación, virtualización en clusters y centros de datos y arquitecturas para plataformas cloud. En el segundo bloque se introduce al alumno en el framework Hadoop para procesamiento Big Data, en cuya arquitectura y utilización se profundizará en las siguientes asignaturas de máster.

1.2 Relación con otras asignaturas

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, sistema operativo Linux 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

Al finalizar la asignatura, el alumno será capaz de:

- Conocer las tecnologías emergentes en el ámbito de los clusters de computación, la virtualización y los centros de datos.
- Comprender los principios de funcionamiento de los clusters implantados sobre centros de datos virtualizados.
- Instalar y administrar una plataforma de virtualización para la computación Big Data.



4. Contenido / Programa de la asignatura

Bloque 1: Infraestructura Cloud para el Big Data

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

En este bloque se aportarán conocimientos básicos sobre las infraestructuras necesarias para el procesamiento Big Data, incluyendo sistemas distribuidos, clusters escalables de computación, virtualización en clusters y centros de datos.

b. Objetivos de aprendizaje

- Conocer las tecnologías emergentes en el ámbito de los clusters de computación, la virtualización y los centros de datos.
- Familiarizarse en el trabajo en plataformas de computación distribuidas desde un enfoque generalista.

c. Contenidos y materiales de aprendizaje

TEMA 1. Distributed Systems, Cluster, Virtualization and Cloud

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, the cloud stack

d. Métodos docentes

- Prácticas guiadas para familiarizarse con entornos de computación distribuida.
- Tutorización a través de foros, email y videoconferencias



e. 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

f. Bibliografía complementaria

- "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
- "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: 978-1-60807-009-8. ISBN ebook: 978-1-60807-010-4

g. Recursos necesarios

- Plataforma docente UbuVirtual
- Máquinas virtuales Ubuntu (serán proporcionadas al alumno)
- Comunicación vía foros y email

h. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO (detallar orden semanas)
0,7	Tema 1.1 -- Semana 1
0,6	Tema 1.2 -- Semana 2
0,6	Tema 1.13-- Semana 3



Bloque 2: Modelos de computación distribuida orientados al Big Data

Carga de trabajo en créditos ECTS:

a. Contextualización y justificación

En este bloque se presenta la infraestructura de Hadoop, uno de los frameworks para procesamiento Big Data más populares y de amplia implantación comercial.

b. Objetivos de aprendizaje

- Conocer la infraestructura del framework Hadoop
- Familiarizarse en el trabajo en la plataforma Cloudera, una distribución de Linux diseñada para Big Data utilizando Hadoop.

c. Contenidos y materiales de aprendizaje

TEMA 2: Introducción a Hadoop y MapReduce

- 2.1 Introducción a Hadoop
- 2.2 Introducción a Modelo de programación MapReduce
- 2.3 Introducción a Cloudera

d. Métodos docentes

- Prácticas guiadas para familiarizarse con Hadoop y la plataforma Cloudera
- Tutorización a través de foros, email y videoconferencias

e. Bibliografía básica

- "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
- "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

f. Bibliografía complementaria

g. Recursos necesarios

- Plataforma docente UbuVirtual
- Máquinas virtuales Ubuntu (serán proporcionadas al alumno)
- Máquina virtual Cloudera (será proporcionada al alumno)
- Comunicación vía foros y email

h. Temporalización

CARGA ECTS	PERIODO PREVISTO DE DESARROLLO (detallar orden semanas)
1	Tema 2 -- Semanas 4-5



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

Actividad Formativa	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. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES	HORAS
Horas de tutoría síncrona o asíncrona	6
Horas de lectura y reproducción materiales	15.5
Horas de trabajo autónomo individual	51.5
Horas de actividades de evaluación	2
Total	75

7. Temporalización (por bloques temáticos)

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Bloque 1	2	Septiembre 2021 (semanas 1-3)
Bloque 2	1	Octubre 2021 (semanas 4-5)

8. Evaluación

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

Crterios / Comentarios a la evaluación

- **Convocatoria extraordinaria:** El alumno deberá realizar las actividades de la asignatura (prácticas o proyecto final) no superadas durante el transcurso de la convocatoria ordinaria. La nota de las actividades ya superadas se mantendrá en la convocatoria extraordinaria. La calificación por participación en los foros de ayuda y debate se obtendrá durante la convocatoria ordinaria, trasladándose la calificación obtenida a la convocatoria extraordinaria.

9. Consideraciones / Comentarios adicionales