

Año académico	2016-17
Asignatura	11638 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 2S
Guía docente	A
Idioma	Castellano

Identificación de la asignatura

Asignatura	11638 - Gestión y Almacenamiento de Datos Masivos
Créditos	1,44 presenciales (36 horas) 4,56 no presenciales (114 horas) 6 totales (150 horas).
Grupo	Grupo 1, 2S (Campus Extens)
Período de impartición	Segundo semestre
Idioma de impartición	Castellano

Profesores

Profesor/a	Horario de atención a los alumnos					
	Hora de inicio	Hora de fin	Día	Fecha inicial	Fecha final	Despacho
Isaac Lera Castro isaac.lera@uib.es	10:00	11:00	Miércoles	26/09/2016	07/07/2017	132
	14:30	16:00	Miércoles	13/02/2017	07/07/2017	132
Pedro Antonio Palmer Rodríguez pere.palmer@uib.es	17:30	18:30	Viernes	01/09/2016	03/03/2017	D142
	11:30	12:30	Viernes	01/09/2016	30/06/2017	D142

Contextualización

Se ha considerado lo siguiente para esta asignatura dentro de la especialización en tecnologías informáticas del MADM.

Con la proliferación de dispositivos y aplicaciones móviles, estamos ante una era con un crecimiento nunca antes visto de movimiento de datos y de diferentes formatos de datos.

Muchas de las empresas no tienen capacidad operativa ni contemplan el abanico de datos disponibles para mejorar la toma de decisiones y mejorar la inteligencia del negocio. Se pierden oportunidades, productividad e ingresos.

En esta asignatura analizaremos las diferentes propuestas en infraestructuras tecnológicas y técnicas que se han realizado. Especialmente, nos centraremos en una de las propuestas con mayor éxito: Apache Hadoop, un proyecto open source que implementa el modelo de programación paralelo propuesto en MapReduce y otras tecnologías de la familia Apache.

Requisitos

No son necesarios requisitos previos. Si es recomendable tener algún conocimiento de shell y/o Java.

Esenciales

No los hay.



Año académico	2016-17
Asignatura	11638 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 2S
Guía docente	A
Idioma	Castellano

Recomendables

No los hay.

Competencias

Las competencias básicas del master se pueden consultar en el siguiente enlace: http://estudis.uib.es/es/master/comp_basiques/

Específicas

- * CE2 Capacidad para la administración y gestión de software para el procesamiento de datos masivos.
- * CE5 Capacidad para la utilización de herramientas disponibles para preparar y ejecutar aplicaciones para datos masivos en la nube..
- * CESP2 Capacidad para seleccionar, atendiendo a criterios de eficiencia, escalabilidad, optimación de acceso, corrección de errores y adecuación al entorno de producción, las bases de datos y el paradigma de datos óptimo en soluciones “Big Data”..

Genéricas

- * CG1 Saber recuperar datos y extraer conocimiento de grandes volúmenes de datos mediante la aplicación eficiente de técnicas de análisis de datos en diferentes dominios. Adoptar los modos de interacción adecuados según las tareas de usuario que se estén apoyando, en especial en aquellos casos en los que interviene el razonamiento analítico..
- * CG4 Comprender y utilizar el lenguaje y las herramientas asociadas al análisis de datos para modelizar y resolver problemas complejos, reconociendo y valorando las situaciones y problemas susceptibles de ser tratados utilizando dichas herramientas y las técnicas asociadas..

Básicas

- * Se pueden consultar las competencias básicas que el estudiante tiene que haber adquirido al finalizar el máster en la siguiente dirección: http://estudis.uib.cat/es/master/comp_basiques/

Contenidos

Contenidos temáticos

1. Perspectiva actual de uso de datos.
Bases de datos estructurales. Requerimientos y Compañías actuales.
Casos de estudio: AWS Amazon y Google Cloud Platform
2. Arquitectura de Sistemas Distribuidos
Modelos de SD/Sistemas Cloud.
Administración. Escalado, Accesibilidad y Complejidad de datos.
Ecosistemas:
Google technologies: GFS, Chubby, B+Tree, Compression Schemes, Bloomfilters, MapReduce
Apache Technologies: Hadoop, HBase, Cassandra, Pig, Hive, ...
3. Introducción a Apache Hadoop



Año académico	2016-17
Asignatura	11638 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 2S
Guía docente	A
Idioma	Castellano

HDFS
Línea de comandos e Interficie Java.
Hadoop I/O files

4. Introducción a Apache Spark
 - Ventajas de SPARK
 - SPARK componentes
 - Creación y uso de un reslient distributed datasets (RDD)
 - Datos con SPARK
5. Monitorización y gestión del cluster
 - Monitorización y Configuración de alarmas
5. MapReduce
 - Anatomía de un programa enMapReduce
 - Streaming en Hadoop
 - Chaining jobs
 - Ejercicios
6. Gestionar MapReduce: gestión, monitorización y pruebas
 - Trabajar con counters
 - Uso de la interface de Hadoop CLI
 - Monitorización de trabajos y su histórico
 - Creación de unit-tests

Metodología docente

Actividades de trabajo presencial

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Clases teóricas	Clases teóricas	Grupo grande (G)	Mediante el método expositivo el profesor establecerá los fundamentos teóricos y prácticos sobre los diferentes aspectos tratados en los temas de la asignatura. Para cada tema se dará información sobre el método de trabajo aconsejable y el material didáctico adicional que el alumno deberá de utilizar para preparar de forma autónoma contenido.	20
Clases prácticas	Clases prácticas	Grupo grande (G)	Los alumnos realizarán trabajos guiados por el profesor donde se mostrará el uso de herramientas de trabajo de la asignatura.	6
Clases de laboratorio	Clases de laboratorio	Grupo mediano (M)	Configuración de entornos Apache	10

Al inicio del semestre estará a disposición de los estudiantes el cronograma de la asignatura a través de la plataforma UIBdigital. Este cronograma incluirá al menos las fechas en las que se realizarán las pruebas de evaluación continua y las fechas de entrega de los trabajos. Asimismo, el profesor o la profesora informará a los estudiantes si el plan de trabajo de la asignatura se realizará a través del cronograma o mediante otra vía, incluida la plataforma Campus Extens.

Año académico	2016-17
Asignatura	11638 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 2S
Guía docente	A
Idioma	Castellano

Actividades de trabajo no presencial

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual o en grupo	Práctica1	Gestión de infraestructura	57
Estudio y trabajo autónomo individual o en grupo	Práctica2	Programación bajo paradigma MapReduce	57

Riesgos específicos y medidas de protección

Las actividades de aprendizaje de esta asignatura no conllevan riesgos específicos para la seguridad y salud de los alumnos y, por tanto, no es necesario adoptar medidas de protección especiales.

Evaluación del aprendizaje del estudiante

Clases prácticas

Modalidad	Clases prácticas
Técnica	Trabajos y proyectos (no recuperable)
Descripción	Los alumnos realizarán trabajos guiados por el profesor donde se mostrará el uso de herramientas de trabajo de la asignatura.
Criterios de evaluación	CG1, CG2, CG4

Porcentaje de la calificación final: 10%

Clases de laboratorio

Modalidad	Clases de laboratorio
Técnica	Trabajos y proyectos (no recuperable)
Descripción	Configuración de entornos Apache
Criterios de evaluación	CG1, CG2, CG4

Porcentaje de la calificación final: 10%



Año académico	2016-17
Asignatura	11638 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 2S
Guía docente	A
Idioma	Castellano

Práctica1

Modalidad	Estudio y trabajo autónomo individual o en grupo
Técnica	Informes o memorias de prácticas (recuperable)
Descripción	Gestión de infraestructura
Criterios de evaluación	CE5, CE2

Porcentaje de la calificación final: 30% con calificación mínima 5

Práctica2

Modalidad	Estudio y trabajo autónomo individual o en grupo
Técnica	Informes o memorias de prácticas (recuperable)
Descripción	Programación bajo paradigma MapReduce
Criterios de evaluación	CESP2,

Porcentaje de la calificación final: 50% con calificación mínima 5

Recursos, bibliografía y documentación complementaria

Transparencias y apuntes por parte de los profesores

Bibliografía básica

Hadoop Apache: <http://hadoop.apache.org/>

Mastering Cloud Computing. Foundations and applications programming. Morgan Kaufman. 2013
Hadoop in Action. Manning. 2010

Bibliografía complementaria

Professional Hadoop Solutions, Wiley. 2013

Hadoop. The definitive Guide. Storage and Analysis at Internet Scale. O'Really Media. 2012

