

Año académico	2016-17
Asignatura	11546 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 1S
Guía docente	A
Idioma	Castellano

Identificación de la asignatura

Asignatura	11546 - Gestión y Almacenamiento de Datos Masivos
Créditos	0,72 presenciales (18 horas) 2,28 no presenciales (57 horas) 3 totales (75 horas).
Grupo	Grupo 1, 1S (Campus Extens)
Período de impartición	Primer 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

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 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.

Requisitos

Competencias

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

Año académico	2016-17
Asignatura	11546 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 1S
Guía docente	A
Idioma	Castellano

Específicas

- * CE7: Capacidad para diseñar, desarrollar, gestionar y evaluar mecanismos de certificación y garantía de seguridad en el tratamiento y acceso a la información en un sistema de procesamiento local o distribuido..
- * CE10: Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería..

Genéricas

- * CG8 - Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contexto más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos..

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. Hadoop
 - HDFS
 - Línea de comandos e Interficie Java.
 - Hadoop I/O files
4. MapReduce - Programming Model
 - Anatomía de un programa enMapReduce
 - Streaming en Hadoop
 - Chaining jobs
 - Ejercicios

Metodología docente

Año académico	2016-17
Asignatura	11546 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 1S
Guía docente	A
Idioma	Castellano

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.	13
Clases prácticas	Clases prácticas	Grupo grande (G)	Los alumnos llevara a término sesiones de trabajo guiadas por el profesor donde se mostrará el uso de herramientas de trabajo de la asignatura.	3
Clases de laboratorio	Clases de laboratorio	Grupo mediano (M)	En el laboratorio se configurará un entorno Hadoop	2

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.

Actividades de trabajo no presencial

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual o en grupo	Práctica 1	Adquisición de conocimientos en la infraestructura Hadoop	22.5
Estudio y trabajo autónomo individual o en grupo	Práctica 2	Adquisición de conocimientos de programación en MapReduce	34.5

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



Año académico	2016-17
Asignatura	11546 - Gestión y Almacenamiento de Datos Masivos
Grupo	Grupo 1, 1S
Guía docente	A
Idioma	Castellano

Práctica 1

Modalidad	Estudio y trabajo autónomo individual o en grupo
Técnica	Trabajos y proyectos (recuperable)
Descripción	Adquisición de conocimientos en la infraestructura Hadoop
Criterios de evaluación	Presentación y ejecución

Porcentaje de la calificación final: 30%

Práctica 2

Modalidad	Estudio y trabajo autónomo individual o en grupo
Técnica	Trabajos y proyectos (recuperable)
Descripción	Adquisición de conocimientos de programación en MapReduce
Criterios de evaluación	Presentación y ejecución

Porcentaje de la calificación final: 70%

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

