



Proyecto docente

Asignatura	Almacenamiento Escalable		
Materia	Tecnologías Informáticas para el Big Data		
Titulación	Máster Universitario en Inteligencia de Negocio y Big Data en Entornos Seguros		
Plan		Código	
Periodo de impartición	Primer Cuatrimestre	Tipo/Carácter	Obligatoria
Nivel/Ciclo	Máster	Curso	2019-2020
Créditos ECTS	3		
Lengua en que se imparte	Castellano		
Profesor/es responsable/s	Aníbal Bregón y Fernando Díaz		
Datos de contacto (e-mail, teléfono...)	Escuela de Ingeniería Informática (Segovia) Plaza de Santa Eulalia 9 y 11, 40005 Segovia Teléfono: 98342300 (ext. 5696) e-mails: anibal@infor.uva.es y fdiaz@infor.uva.es		
Horario de tutorías	Disponible en http://www.inf5g.uva.es/?q=node/20		
Coordinador			
Departamento	Informática (ATC, CCIA, LSI)		



1. Situación / Sentido de la asignatura

1.1 Contextualización

La asignatura Almacenamiento Escalable se encuadra dentro de la materia Tecnologías Informáticas para el Big Data y ofrece al alumno los conocimientos fundamentales para entender el reto que supone preservar grandes colecciones de datos y las tecnologías más destacadas que existen para abordar dicho reto.

La escalabilidad del almacenamiento de Big Data está directamente relacionada con la utilización de sistemas de archivos distribuidos. Este tipo de infraestructura permite añadir recursos de almacenamiento con los que afrontar las necesidades crecientes que presentan los sistemas informáticos que gestionan grandes colecciones de datos. En la actualidad, el sistema de archivos HDFS (*Hadoop Distributed File System*) es la solución más utilizada en el ámbito Big Data y, por tanto, la referencia a manejar en este ámbito. Los recursos que proporciona HDFS han servido para el desarrollo de soluciones de almacenamiento de más alto nivel: las bases de datos NoSQL (*Not Only SQL*). Aunque existen múltiples tipos de soluciones NoSQL, todas ellas comparten una naturaleza distribuida y que, por tanto, garantiza su escalabilidad. A lo largo de esta asignatura se profundizará en HDFS y en las bases de datos NoSQL más utilizadas en diferentes áreas centradas en la explotación de Big Data.

En resumen, esta asignatura se divide en dos bloques temáticos diseñados para que el alumno obtenga los conocimientos necesarios para poder tomar decisiones efectivas de almacenamiento de Big Data. En el primer bloque se introducirán los conceptos principales sobre sistemas de ficheros distribuidos y se presentará HDFS tanto a nivel teórico como práctico. En el segundo bloque se abordarán los principios fundamentales de la tecnología NoSQL y se presentarán algunos de los sistemas de bases no relacionales más destacados en el ámbito del Big Data.

1.2 Relación con otras asignaturas

El almacenamiento de Big Data es un aspecto transversal a cualquier sistema informático que gestione grandes colecciones de datos. Por lo tanto, los contenidos impartidos en esta asignatura están relacionados de forma directa con otras asignaturas del plan de estudios, en particular con Arquitecturas Big Data e Infraestructura para el Big Data.

1.3 Prerrequisitos

Se recomienda que el alumno, en sus estudios de grado, haya adquirido un mínimo de competencias en relación con el uso, configuración y administración, y conocimiento de los lenguajes de programación utilizados en sistemas operativos, sistemas distribuidos y sistemas de bases de datos.



2. Competencias

2.1 Generales del título

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

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 ...

- Entender como el uso de sistemas de ficheros distribuidos es aplicable al Big Data, cómo almacenar y consultar Big Data utilizando los entornos actualmente disponibles.
- Aplicar bases de datos no relacionales, las técnicas para almacenar grandes volúmenes de datos estructurados y no estructurados.
- Ser capaz de entender los retos que supone el almacenamiento de Big Data y como todos ellos pasan por utilizar técnicas de distribución de datos.
- Ser capaz de comprender los principios fundamentales de los sistemas de ficheros distribuidos y ponerlos en práctica con HDFS.
- Ser capaz de entender las capacidades específicas de los modelos principales de almacenamiento NoSQL y sus diferencias respecto al modelo relacional.
- Ser capaz de identificar un problema Big Data y elegir el mejor modelo de almacenamiento NoSQL para afrontarlo.
- Ser capaz de utilizar algunas de las bases de datos NoSQL más demandadas en los escenarios Big Data actuales.



4. Contenido / Programa de la asignatura

4.1 Unidades docentes (bloques de contenidos)

- Introducción: sistemas de ficheros distribuidos, requisitos del almacenamiento escalable.
- HDFS (Hadoop Distributed File System): diseño y arquitectura de HDFS, conceptos fundamentales de HDFS, interacción con HDFS, Dataflow, Blocks y Replicación.
- Procesos Hadoop: Name node, Secondary name node, Job tracker, Task tracker, Data node. Administración, monitorización y mantenimiento HDFS
- Almacenamiento en BBDD NoSQL: introducción, conceptos fundamentales, principales tipos de BBDD NoSQL.
- Almacenamiento clave-valor: introducción, ideas principales, HBase, Cassandra.
- Almacenamiento documental: introducción, ideas principales.
- Otras aproximaciones de almacenamiento NoSQL.

4.2 Bibliografía

WHITE, T. "Hadoop: The Definitive Guide". 4th Ed. O'Reilly Media. 2015.

TIWARI, SHASHANK, "Professional NoSQL", Wiley/Wrox. 2011.

STRAUCH, CHRISTOF, "NoSQL Databases". Stuttgart Media University. 2012.

CHODOROW, K. "MongoDB: The Definitive Guide", 2nd Edition, O'Reilly Media. 2013.

CARPENTER, J., HEWITT, E. "Cassandra: The Definitive Guide", 2nd Edition, O'Reilly Media. 2016.

VAISH, G., "Getting Started with NoSQL". Packt Publishing. 2013.



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

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



6. Temporalización (por bloques temáticos)

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
Introducción	0,2	Noviembre 2018
HDFS	0,4	Noviembre 2018
Procesos Hadoop	0,4	Diciembre 2018
Almacenamiento NoSQL	2	Diciembre 2018



7. Evaluación

Instrumento / Procedimiento	Peso primera convocatoria	Peso segunda convocatoria
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%



8. Recursos de aprendizaje y apoyo tutorial del curso online

Transparencias.
Enunciados de ejercicios.
Cuestionarios de autoevaluación.
Páginas Webs relacionadas
Bibliografía disponible en la Biblioteca
Tutorías individualizadas o en grupo a demanda de los alumnos.



9. Consideraciones / Comentarios adicionales
