

DATOS GENERALES DE LA ASIGNATURA

ASIGNATURA: Tecnologías de Almacenamiento de Datos

PLAN DE ESTUDIOS: Máster Universitario en Gestión y Análisis de Grandes Volúmenes de Datos: Big Data

GRUPO: 2223-O1

CENTRO: Escuela Politécnica Superior

CARÁCTER DE LA ASIGNATURA: Obligatorio

ECTS: 6,0

CURSO: 1º

SEMESTRE: 1º Semestre

IDIOMA EN QUE SE IMPARTE:

Castellano, excepto las asignaturas de idiomas que se impartirán en el idioma correspondiente

DATOS DEL PROFESOR

NOMBRE Y APELLIDOS: FERNANDO ALONSO PASTOR

EMAIL: falonsop@uemc.es

TELÉFONO: 983 00 10 00

CV DOCENTE:

Ingeniero Informático por la Universidad de Valladolid.
Máster en Ingeniería Informática con especialización en Inteligencia Artificial por la Universidad Politécnica de Madrid.
Actualmente cursando los estudios de Doctorado en Software, Sistemas y Computación en la Universidad Politécnica de Madrid.
Tutor de TFM en el Máster Universitario de Big Data en la Universidad Europea Miguel de Cervantes.

Ingeniero Informático por la Universidad de Valladolid.
Máster en Ingeniería Informática con especialización en Inteligencia Artificial por la Universidad Politécnica de Madrid.
Actualmente cursando los estudios de Doctorado en Software, Sistemas y Computación en la Universidad Politécnica de Madrid.
Tutor de TFM en el Máster Universitario de Big Data en la Universidad Europea Miguel de Cervantes.

CV PROFESIONAL:

Data Scientist en Accenture, desde Septiembre 2019
Software Engineer en HP SCDS, 2019
Investigador en grupo Trasgo (MoBiVAP) en Universidad de Valladolid, 2018-2019

CV INVESTIGACIÓN:

Investigador colaborador en el Grupo de investigación en Información y Computación Cuántica (GIICC) en la Universidad Politécnica de Madrid, desde 2022.
Investigador en grupo Trasgo (MoBiVAP) en Universidad de Valladolid, 2018-2019.

DATOS ESPECÍFICOS DE LA ASIGNATURA

DESCRIPTOR DE LA ASIGNATURA:

Durante el desarrollo de la asignatura de TECNOLOGÍAS PARA EL ALMACENAMIENTO DE DATOS, realizaremos un recorrido por algunas de las herramientas que se sitúan en el ecosistema Big Data.

Comenzaremos con una introducción al mundo del Big Data, su definición y alguna de las aplicaciones que este famoso término tiene el mundo real. Big Data es un término en auge, pero también es un término ‘virgen’, del cual aún es difícil dar una definición clara y acotar a que llamamos con este nombre. Estudiaremos cuales son las claves y los conceptos básicos en los que se basa el Big Data para funcionar.

Cuando se almacenan grandes volúmenes de datos de forma distribuida los esquemas más trabajados hasta el momento, conocidos como esquemas relacionales, dejan de ser útiles por la dispersión de los datos, la ausencia de información estructurada, la dificultad en representar las complejas interrelaciones que existen entre los datos, el tiempo de respuesta requerido o una conjunción de estos factores. Es por este motivo que surgen nuevos paradigmas fuera de las convenciones actuales, a los que se conoce como bases de datos NoSQL.

Esta asignatura nos permite poner los pilares en el tratamiento de Big Data, ya que permite disponer de mecanismos para representar los datos no estructurados, grabarlos, consultarlos, modificarlos y eliminarlos de acuerdo a necesidades específicas de dominios donde se gestiona mucha información y se deben tomar decisiones en base a la misma.

El objetivo fundamental de este módulo es alcanzar una visión global de qué se entiende por una base de datos NoSQL, proporcionando al estudiante una visión de los múltiples conceptos, modelos y herramientas que se pueden agrupar bajo esta denominación.

El alumno no es necesario que tenga conocimientos previos acerca del mundo de Big Data, aunque sí es recomendable tener conocimientos acerca de los lenguajes de programación sql y scala. Por último, sería también recomendable tener ciertos conocimientos sobre comandos Linux.

CONTENIDOS DE LA ASIGNATURA:

1. **Introducción a las tecnologías de almacenamiento de datos.**
 1. Conceptos: Introducción NoSQL, persistencia polígota, teorema CAP, introducción a los modelos de agregación, introducción al modelo de grafo.
2. **Modelos de agregación: Bases de datos documentales: MongoDB**
 1. Conceptos: Introducción al modelo documental, consideraciones de diseño en el modelo documental, modelo de datos de MongoDB, introducción a MongoDB, Operaciones CRUD en MongoDB, distribución de datos en MongoDB
3. **Modelos de agregación: Bases de datos por columnas: Cassandra**
 1. Conceptos: Introducción al modelo de columnas, consideraciones de diseño en el modelo de columnas, introducción a Cassandra, operaciones CRUD en Cassandra, distribución de datos en Cassandra.
4. **Almacenamiento y procesamiento distribuido: Apache Hadoop**
 1. Conceptos: Introducción a la arquitectura de Hadoop, HDFS y MapReduce
5. **Procesamiento Distribuido: Apache Spark**
 1. Conceptos: introducción a Apache Spark, modelo de grafos adirigidos, acciones y transformaciones, introducción a los SDKs de Spark. Conceptos de Streaming, introducción al streaming con Spark

RECURSOS DE APRENDIZAJE:

Los recursos de aprendizaje que se utilizarán en todas las asignaturas de la titulación (salvo las prácticas externas) para facilitar el proceso de enseñanza-aprendizaje, son:

- Campus online de la UEMC (Open Campus)
- Plataforma de Webconference (Adobe Connect)

Las comunicaciones con el profesor serán a través de Open Campus vía Mi correo, Tablón o/y Foro.

COMPETENCIAS BÁSICAS:

- CB6. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- CB7. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- CB8. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
- CB9. Que los estudiantes sepan comunicar sus conclusiones -y los conocimientos y razones últimas que las sustentan- a públicos especializados y no especializados de un modo claro y sin ambigüedades
- CB10. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

COMPETENCIAS GENERALES:

- CG01. Capacidad de organización y planificación
- CG02. Comunicación oral y escrita en lengua extranjera
- CG03. Habilidades básicas de informática
- CG04. Capacidad y habilidad para la toma de decisiones
- CG05. Capacidad para trabajar en equipos de carácter interdisciplinar
- CG06. Compromiso ético (saber aplicar la evidencia científica en la práctica profesional y mantener un compromiso ético y de integridad intelectual en el planteamiento de la investigación científica, básica y aplicada)
- CG07. Capacidad de crítica y autocrítica
- CG08. Habilidades interpersonales (tanto con miembros del entorno como con científicos/profesionales de otros centros)
- CG09. Reconocimiento a la diversidad y a la multiculturalidad
- CG10. Desarrollar hábitos de excelencia y calidad en el ejercicio profesional
- CG11. Capacidad para generar nuevas ideas (creatividad)

COMPETENCIAS ESPECÍFICAS:

- CE11. Capacidad para definir la arquitectura de una plataforma BigData y reconocer los servicios que conforman el ecosistema de una plataforma
- CE12. Capacidad para determinar el volumen de almacenamiento, memoria y capacidad de procesamiento en la plataforma para realizar un correcto dimensionamiento de los nodos que componen el cluster Big Data.

RESULTADOS DE APRENDIZAJE:

El alumno será capaz de:

- El alumno será capaz de utilizar las técnicas y herramientas básicas para almacenar y extraer información estructurada en bases de datos relacionales
- El alumno será capaz de reconocer los servicios que conforman la arquitectura de una plataforma de grandes volúmenes de datos

BIBLIOGRAFÍA Y RECURSOS DE REFERENCIA GENERALES

BIBLIOGRAFÍA BÁSICA:

- Rafael Caballero, Enrique Martín (2015): Las bases del Big Data. Catarata. ISBN: 978-8490970867

BIBLIOGRAFÍA COMPLEMENTARIA:

- Boris Lublinsky , Kevin T.Smith, Alexey Yakubovich (2014): Hadoop. Soluciones Big Data. Anaya. ISBN: 978-

8441535916

- Bill Chambers, Matei Zaharia (2018): Spark: The Definitive Guide: Big Data Processing Made Simple. O'Reilly Media. ISBN: 978-1491912218

WEBS DE REFERENCIA:

Web / Descripción

[Apache Hadoop](http://hadoop.apache.org/docs/stable/)(http://hadoop.apache.org/docs/stable/)
Web de Apache Hadoop

[Apache Cassandra](http://cassandra.apache.org/doc/latest/)(http://cassandra.apache.org/doc/latest/)
Web de Apache Cassandra

[Apache Spark](http://spark.apache.org/)(http://spark.apache.org/)
Web de Apache Spark

[Spark SQL](https://spark.apache.org/docs/latest/sql-programming-guide.html)(https://spark.apache.org/docs/latest/sql-programming-guide.html)
Documentación SparkSQL

[Mongo DB](https://docs.mongodb.com/manual/reference/database-references/)(https://docs.mongodb.com/manual/reference/database-references/)
Documentación de referencia de MongoDB

PLANIFICACIÓN DEL PROCESO DE ENSEÑANZA-APRENDIZAJE DE LA ASIGNATURA

METODOLOGÍAS:

Se describe a continuación la metodología aplicada

MÉTODO DIDÁCTICO:

El papel del profesor cobra importancia a través de la impartición de clases magistrales en tiempo real por videoconferencia que podrá utilizar para explicar los contenidos teóricos, resolver dudas que se planteen durante la sesión, ofrecer retroalimentación sobre las actividades de evaluación continua o realizar sesiones de tutoría de carácter grupal.

MÉTODO DIALÉCTICO:

Se caracteriza por la participación de los alumnos en las actividades de evaluación continua de debate y la intervención de éstos a través del diálogo y de la discusión crítica (seminarios, grupos de trabajo, etc.). Utilizando este método el alumno adquiere conocimiento mediante la confrontación de opiniones y puntos de vista. El papel del profesor consiste en proponer a través de Open Campus temas referidos a la materia objeto de estudio que son sometidos a debate para, posteriormente, evaluar el grado de comprensión que han alcanzado los alumnos.

MÉTODO HEURÍSTICO:

Este método puede desarrollarse de forma individual o en grupo a través de las actividades de evaluación continua (entregas de trabajos, resolución de ejercicios, presentaciones, etc.). El objetivo es que el alumno asuma un papel activo en el proceso de aprendizaje adquiriendo los conocimientos mediante la experimentación y la resolución de problemas.

CONSIDERACIONES DE LA PLANIFICACIÓN:

Las actividades formativas que se realizan en la asignatura son las siguientes:

Clases teóricas: Actividad dirigida por el profesor que se desarrollará de forma sincrónica en grupo. Para la realización de esta actividad en Open Campus, la UEMC dispone de herramientas de Webconference que permiten una comunicación unidireccional en las que el docente puede desarrollar sesiones en tiempo real con posibilidad de ser grabadas para ser emitidas en diferido.

Actividades prácticas: Actividades supervisadas por el profesor que se desarrollarán fundamentalmente de forma asíncrona, y de forma individual o en grupo:

- Actividades de debate. Se trata de actividades en las que se genera conocimiento mediante la participación de los estudiantes en discusiones alrededor de temas de interés en las distintas

asignaturas.

- o Entregas de trabajos individuales o en grupos a partir de un enunciado o unas pautas de trabajo que establecerá el profesor.
- o Resolución de ejercicios y problemas que el alumno debe realizar a través de Open Campus en un periodo de tiempo determinado. Esta actividad puede ser en formato test de evaluación.

Tutorías: Las tutorías podrán tener un carácter sincrónico o asíncrono y podrán desarrollarse de manera individual o en grupos reducidos.

Están previstas dos sesiones de tutoría por videoconferencia, una al inicio y otra al final del semestre. En la primera se presentará la asignatura y la guía docente y en la segunda, en las semanas previas a la evaluación final, se dedicará a la resolución de dudas de los estudiantes.

Además, el docente utiliza el Tablón, el Foro y el Sistema de correo interno de Open Campus para atender las necesidades y dudas académicas de los estudiantes.

SESIONES EN TIEMPO REAL :

	Título
TU1	Presentación asignatura y Guía docente
CM1	Introducción a las tecnologías de almacenamiento de datos
CM2	MongoDB
CM3	Presentación de las Prácticas. Resolución de dudas.
CM4	Cassandra
CM5	Hadoop
CM6	Spark
CM7	Spark Streaming
TU2	Resolución de dudas antes de la evaluación

EVALUACIÓN CONVOCATORIA ORDINARIA:

Evaluación continua	60%
Evaluación final	40%

ACTIVIDADES Y SISTEMAS DE EVALUACIÓN :

Tipo Evaluación	Nombre Actividad	% Calif.
Evaluación continua (60 %)	1. Actividad 1. Entrega individual (Entrega individual)	25
	2. Actividad 2. Entrega individual (Entrega individual)	25
	3. Actividad 3. Test evaluación (Test de evaluación)	10
Evaluación final (40 %)	1. Evaluación final convocatoria ordinaria (Prueba de evaluación final presencial)	40

CONSIDERACIONES EVALUACIÓN CONVOCATORIA ORDINARIA:

A lo largo de la planificación de la asignatura el alumno realizará **actividades de evaluación continua** que forman parte de la calificación de la asignatura con un peso del 60% sobre la nota final.

Para superar la evaluación continua, el alumno debe obtener una media de igual o superior a 5 entre todas las actividades. En el caso de no superar la evaluación continua, se guardan para la convocatoria extraordinaria las notas de aquellas actividades aprobadas, no pudiendo volver a presentarlas.

El sistema de evaluación de esta asignatura acentúa el desarrollo gradual de competencias y resultados de aprendizaje y, por tanto, se realizará una evaluación continua a través de las distintas actividades de evaluación propuestas. El resultado de la evaluación continua se calcula a partir de las notas obtenidas en cada actividad teniendo en cuenta el porcentaje de representatividad en cada caso.

Todas las actividades deberán entregarse en las fechas previstas para ello, teniendo en cuenta:

- Las actividades de evaluación continua (entrega de trabajos) se desarrollarán según se indica y, para ser evaluadas, los trabajos deberán ser entregados en la forma y fecha prevista y con la extensión máxima señalada. No se evaluarán trabajos entregados posteriormente a esta fecha o que no cumplan con los criterios establecidos por el profesor.
- La no entrega de una actividad de evaluación continua en forma y plazo se calificará con un 0 y así computarán en el cálculo de la nota de evaluación continua y final de la asignatura.
- Cualquier tipo de copia o plagio por mínimo que sea, supondrá una calificación de 0 en la actividad correspondiente.
- Las actividades de evaluación continua (tipo test) se desarrollarán con anterioridad a la realización de las pruebas de evaluación parcial y final de la asignatura

Los alumnos accederán a través de OpenCampus a las calificaciones de las actividades de evaluación continua en un plazo no superior a 15 días lectivos desde su fecha de entrega, excepto causas de fuerza mayor en cuyo caso se informará al alumno a través del Tablón.

La evaluación continua se completará con una **evaluación final online**, que se realizará al finalizar el periodo lectivo de cada asignatura. Los exámenes serán eminentemente prácticas, de manera que, los alumnos podrán disponer de los apuntes y consultarlos, (solo en formato digital) durante la realización de la prueba.

Para resolver el examen, los alumnos deberán descargar el enunciado de la prueba y una vez cumplimentado, subirlo en el espacio correspondiente del campus virtual

La prueba **supondrá un 40%** de la calificación sobre la nota final de la asignatura.

- El alumno tendrá la posibilidad, siempre dentro de los tres días siguientes a la publicación de las notas, a renunciar a su calificación, y presentarse en la siguiente convocatoria
- El alumno tendrá hasta 3 días después de la calificación para solicitar al docente más información sobre su calificación por el correo de la plataforma.ÍA DOCENTE
- Cualquier tipo de irregularidad o fraude en la realización de una prueba, supondrá una calificación de 0 en la prueba/convocatoria correspondiente.
- El aplazamiento concedido por la Universidad para la realización de una evaluación final se registrará por lo establecido en el Manual de "Directrices y plazos para la tramitación de una solicitud"

La nota final se corresponderá con la media aritmética del resultado obtenido en cada una de las partes. En caso de no superación, se guarda la parte aprobada para la convocatoria extraordinaria.

La **nota global** de la asignatura se obtiene ponderando la calificación de la evaluación continua y de la evaluación final según los siguientes porcentajes, y debiendo tener aprobadas ambas partes, continua y final, para superar la asignatura.

Si un alumno no se presenta a la prueba de evaluación final, su calificación en la convocatoria será de "No presentado", con independencia de que haya realizado alguna actividad de evaluación continua.

De igual modo si el alumno no entrega ninguna actividad de evaluación continua, obtendrá la calificación de "No presentado", con independencia de que haya aprobado la prueba de evaluación final, en cuyo caso, se le guardaría su calificación para la convocatoria extraordinaria

EVALUACIÓN CONVOCATORIA EXTRAORDINARIA:

Evaluación continua	60%
Evaluación final	40%

ACTIVIDADES Y SISTEMAS DE EVALUACIÓN :

Tipo Evaluación	Nombre Actividad	% Calif.
Evaluación continua (60 %)	1. Actividad 1. Entrega individual (Entrega individual)	25
	2. Actividad 2. Entrega individual (Entrega individual)	25
	3. Actividad 3. Test evaluación (Test de evaluación)	10
Evaluación final (40 %)	1. Evaluación final convocatoria extraordinaria (Prueba de evaluación final presencial)	40

CONSIDERACIONES EVALUACIÓN CONVOCATORIA EXTRAORDINARIA:

Los estudiantes que no hayan superado la asignatura en la convocatoria ordinaria, porque hayan suspendido la evaluación continua o la prueba de evaluación final, podrán presentarse a las pruebas establecidas por el profesor en la convocatoria extraordinaria.

Para la convocatoria extraordinaria se guardan las calificaciones de las actividades de evaluación continua y prueba de evaluación final, superadas por el estudiante (nota superior o igual a 5), no permitiéndose volver a realizarlas.

- En convocatoria extraordinaria, el alumno solo podrá entregar las actividades de evaluación continua no superadas, guardándose la calificación de las aprobadas.
- El alumno tendrá hasta 3 días después de la calificación para solicitar al docente más información sobre su calificación por el correo de la plataforma.
- Cualquier tipo de irregularidad o fraude en la realización de una prueba, supondrá una calificación de 0 en la prueba/convocatoria correspondiente.
- El aplazamiento concedido por la Universidad para la realización de una evaluación final se registrará por lo establecido en el Manual de "Directrices y plazos para la tramitación de una solicitud".

En la convocatoria extraordinaria, la **nota global** de la asignatura se obtiene ponderando la calificación de la evaluación continua y de la evaluación final, de la misma forma que en la convocatoria ordinaria.

Al igual que en la convocatoria ordinaria, en la convocatoria extraordinaria es necesario superar tanto la evaluación continua como la evaluación final para aprobar la asignatura.

Si un alumno no se presenta a la prueba de evaluación final, su calificación en la convocatoria será de "No presentado", con independencia de que haya realizado alguna actividad de evaluación continua.

SISTEMAS DE EVALUACIÓN:

SISTEMA DE EVALUACIÓN	PORCENTAJE (%)
Pruebas de ejecución de tareas reales y/o simuladas	10%
Pruebas de respuesta corta	20%
Pruebas de respuesta larga, de desarrollo	30%
Trabajos y proyectos	40%