



FICHA TÉCNICA DE LA ASIGNATURA

| Datos de la asignatura | |
|------------------------|---|
| Nombre completo | Bases de Datos NoSQL |
| Código | DTC-MBD-528 |
| Título | N/A |
| Impartido en | Máster Universitario en Big Data [Primer Curso] |
| Créditos | 6,0 ECTS |
| Carácter | Obligatoria |
| Departamento / Área | Departamento de Telemática y Computación |

| Datos del profesorado | |
|-----------------------|--|
| Profesor | |
| Nombre | Enrique Fernández-Pampillón Cesteros |
| Departamento / Área | Departamento de Telemática y Computación |
| Correo electrónico | efernandezpampillon@icai.comillas.edu |
| Profesor | |
| Nombre | Diego Yagüe Juárez |
| Departamento / Área | Departamento de Telemática y Computación |
| Correo electrónico | dyague@icai.comillas.edu |

DATOS ESPECÍFICOS DE LA ASIGNATURA

| Contextualización de la asignatura |
|--|
| Aportación al perfil profesional de la titulación |
| <p>Hoy en día el paradigma de los datos ha cambiado el mundo, e incluso la economía se ve afectada por los datos; una empresa no sólo vale lo que produce sino los datos que tiene.</p> <p>La cantidad de datos que se producen y procesan en todo el mundo ha ido creciendo en los últimos años a medida que los dispositivos móviles y muchos otros, como cámaras de vigilancia, monitores y sistemas de medición, han ido evolucionando a tener una conexión a internet; pero no sólo las máquinas son las responsables de este ingente incremento, las interacciones humanas, es decir, las redes sociales, reseñas de servicios/productos, blogs, etc., se han convertido en una fuente de datos muy importante para que las empresas o las partes interesadas extraigan valor de ellas.</p> <p>En este escenario, los sistemas de almacenamiento se han vuelto imprescindibles, se precisan de sistemas con capacidad para gestionar grandes volúmenes de datos (del orden de Petabytes) y con un alto rendimiento. Y en esta encrucijada, los sistemas tradicionales no son siempre la solución adecuada.</p> |



Actualmente, los casos de uso de BigData no son posibles sin una base de datos NoSQL y teniendo en cuenta, las tendencias con las comunicaciones 5G, el aumento de aparatos electrónicos o "inteligentes", y las necesidades de seguridad de los datos, el conocimiento, gobernanza y gestión de los datos se torna en imprescindible en las empresas.

Como Ingeniero de Datos, Analista de Datos o Científico de Datos, es obligatorio tener algunas habilidades en Bases de Datos NOSQL para aprovechar sus conocimientos y diseñar Soluciones de Negocio.

| Competencias - Objetivos | |
|-----------------------------------|---|
| Competencias | |
| Conocimientos o contenidos | |
| CO1 | Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinarias para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área. |
| CO2 | Comprender las técnicas de procesados de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos. |
| Competencias | |
| CP1 | Integrar las arquitecturas, técnicas de inteligencia artificial, análisis avanzado de datos y de visualización y de cumplimiento legal para ofrecer la solución global óptima. |
| CP2 | Aplicar e integrar los flujos programáticos de datos masivos |
| CP4 | Implementar las técnicas de procesamiento de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos. |
| CP7 | Aplicar conocimientos avanzados en Big Data y analítica de datos para desarrollar soluciones innovadoras en proyectos y en investigación, aportando y evaluando soluciones óptimas para el procesamiento y análisis de datos a gran escala. |
| Habilidades o destrezas | |
| HA1 | Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos. |
| HA2 | Trabajar en equipos de carácter pluridisciplinar y/o internacional y organizar y liderar adecuadamente las dinámicas de grupo. |
| HA3 | Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos). |
| HA4 | Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad. |
| HA5 | Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos. |



BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Introducción a las bases de datos SQL y No SQL

Introducción a las bases de datos SQL y No SQL

Historia de las BBDD.

Transaccionabilidad: ACID y BASE, Teorema de Brewer.

Arquitectura de las Base de Datos -Subsistemas y Funcionalidades.

Tipologías de Bases de Datos: Columnares, Clave-Valor, Documentales, In-Memory, etc

Bases de datos relacionales.

Bases de datos relacionales

Diseño y normalización.

Gestión de bases de datos.

Arquitectura de BBDD Relacionales. Beneficios e inconvenientes.

Modelado de datos.

Normalización de los modelos.

Modelados en Big-Data. Modelos en Estrella. Dimensiones y Hechos.

Procesos de carga masiva de datos – ETL.

Estructuras y Best Practices para ETL.

Bases de datos NoSQL Clave-Valor.

Bases de datos NoSQL Clave-Valor

Introducción.

Beneficios e inconvenientes. Arquitectura– Componentes y Regiones.

Sistema de Almacenamiento.

Diseño de Claves.

Modelados.

Operaciones DDL y DML.

Operaciones de recuperación de información. Filtros.

Procesos de carga masiva. Compactación

Bases de datos NoSQL de gestión de información basada en documentos.



Bases de datos NoSQL de gestión de información basada en documentos

Introducción a las bases de datos documentales. Beneficios e inconvenientes.

Arquitectura de Componentes.

Arquitectura de Replicas.

Análisis sistema de Almacenamiento. Modelado.

Estructuras de índices para acelerar la explotación de datos

Operaciones DDL y DML. • Map-Reduce y Proyecciones.

Bases de Datos NoSQL de gestión de grafos.

Bases de Datos NoSQL de gestión de grafos

El paradigma de los grafos.

Datos vs Grafos. Beneficios e inconvenientes.

Conceptos de Grafos: Nodos y relaciones.

Modelado.

Operaciones de inserción y recuperación.

Anidación, Índices y constraints.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

| | |
|---|------------------------------|
| Clases magistrales expositivas y participativas | CO1, CO2, CP1, HA1 |
| Ejercicios prácticos y resolución de problemas | CP1, CP4, CP7, HA1, HA5, CP2 |

Metodología No presencial: Actividades

| | |
|--|---|
| Ejercicios prácticos y resolución de problemas | CP1, CP4, CP7, HA1, HA4, HA5, CP2 |
| Estudio personal | CO1, CO2, CP1, HA4, HA5 |
| Proyectos y Trabajos | CP1, CP4, CP7, HA1, HA2, HA3, HA4, HA5, CP2 |

RESUMEN HORAS DE TRABAJO DEL ALUMNO



| HORAS PRESENCIALES | | | |
|--|---|---|---|
| Clases magistrales expositivas y participativas: Exposición de contenidos fundamentales por parte del profesor impulsando la reflexión y participación del alumno. | Ejercicios prácticos y resolución de problemas: Sesiones prácticas con uso de software: Actividad formativa con ordenador que, bajo la guía del profesor-tutor, fomenta el aprendizaje autónomo y/o cooperativo del alumno mediante la ejecución de programas para la consecución de los objetivos marcados | | |
| 30.00 | 30.00 | | |
| HORAS NO PRESENCIALES | | | |
| Estudio personal: Reflexión y análisis individual de los contenidos teóricos y prácticos de las materias y/o asignaturas del Master | Trabajos: Los alumnos tendrán que hacer trabajos breves (individuales y/o en grupo), por indicación del profesor | Proyectos: Los alumnos tendrán que hacer trabajos de tamaño medio o grande (individuales y/o en grupo), por indicación del profesor | Ejercicios prácticos y resolución de problemas: Sesiones prácticas con uso de software: Actividad formativa con ordenador que, bajo la guía del profesor-tutor, fomenta el aprendizaje autónomo y/o cooperativo del alumno mediante la ejecución de programas para la consecución de los objetivos marcados |
| 30.00 | 30.00 | 30.00 | 30.00 |
| CRÉDITOS ECTS: 6,0 (180,00 horas) | | | |

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

| Actividades de evaluación | Criterios de evaluación | Peso |
|------------------------------------|--|------|
| Examen Final | <ul style="list-style-type: none"> • Comprensión de los contenidos teóricos • Aplicación de los contenidos a resolución de problemas • Análisis crítico de resultados | 70 % |
| Trabajos prácticos de Laboratorio. | <ul style="list-style-type: none"> • Aplicación de los contenidos teóricos para la resolución de problemas reales. • Habilidades de comunicación escrita. • Habilidades de comunicación Oral: Exposición oral ejecutiva de la práctica: Diseño de solución, Resultados, Inconvenientes, Lecciones Aprendidas. | 20 % |
| Trabajo Individual | <ul style="list-style-type: none"> • Participación Individual mostrando la aplicación de los contenidos teóricos para la resolución de problemas reales. • Habilidades de comunicación escrita y oral tanto en su aportación a los resultados del grupo como en sus participaciones individuales. | 10 % |

Calificaciones

La nota final del curso será calculada en base a la siguiente estructura:

Examen Final 70% Será necesario un resultado mínimo del 40% del examen para poder aprobar la asignatura.
Prácticas de
Laboratorio 20%
Participación en
clase 10%

PLAN DE TRABAJO Y CRONOGRAMA

| Actividades | Fecha de realización | Fecha de entrega |
|--|--|------------------|
| | S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 | |
| Introducción a las bases de datos SQL y No SQL: | Teoría P P | |
| Bases de datos relacionales. | Teoría P P | |
| Laboratorio: Modelo y Gestión de Información Relacional | Práctica A A/P A A A | |
| Bases de datos NoSQL Clave-Valor. | Teoría P P P | |
| Laboratorio: Modelado y Gestión Información en Clave-Valor | Práctica A A/P A A A | |
| Bases de datos NoSQL de gestión de información basada en documentos. | Teoría P P P | |
| Laboratorio: Modelado y Gestión | Práctica A A A A | |



| | | | |
|---------------------------|--|---|---|
| Documentos | | | |
| Bases de Datos | | | |
| NoSQL de Teoría | | P | P |
| gestión de grafos. | | | |
| Laboratorio: | | | |
| Manejo de Práctica Grafos | | | P |
| A -> Autoestudio | | | |
| P -> Presencial | | | |

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

- Bibliografía Básica Slides prepared by the lecturer (available in Moodlerooms)
- Shripav , Shashwat - Learning HBase .Birmingham, Packt Publishing.
- George, Lars (2011): HBase : The Definitive Guide. Sebastopol,O'Reilly
- Banker , Kyle, Peter Bakkum , Shaun Verch ,Douglas Garrett, Tim Hawkins (2016): MongoDB in Action . Second Edition . New York, Manning Publications Co.
- Vukotic , Aleksa , Nicki Watt (2015): Neo4j in Action . New York, Manning Publications

Bibliografía Complementaria

- Spaggiari, Jean-Marc, Kevin O'Dell (2016): Architecting HBase Applications . Sebastopol,O'Reilly.
- <https://www.postgresql.org/docs/>
- https://hbase.apache.org/apache_hbase_reference_guide.pdf
- <https://hbase.apache.org/book.html#quickstart>
- <https://www.mongodb.com/docs/>
- Panzarino , Onofrio (2014): Learning Cypher. Birmingham , Packt Publishing
- <https://neo4j.com/docs>