
Curso Académico: (2024 / 2025)**Fecha de revisión: 22-05-2024**

Departamento asignado a la asignatura: Departamento de Informática**Coordinador/a: CALLE GOMEZ, FRANCISCO JAVIER****Tipo: Obligatoria Créditos ECTS : 6.0****Curso : 1 Cuatrimestre : 1**

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

- Bases de Datos Estructuradas
- Lenguajes de Datos Algebraicos
- SQL
- Bases de Datos OLAP y Data Warehouse

Habilidades de programación (conveniente fundamentos de Javascript)

OBJETIVOS

- Comprender el concepto y todas las dimensiones de la tecnología Big Data
- Explorar los contextos social, empresarial y tecnológico que subyacen al surgimiento y expansión de esta tecnología
- Comprender el Ciclo de Vida de la Información, y los procesos que lo sostienen
- Analizar las necesidades que suscita la Información: adquisición, transformación, almacenamiento y explotación
- Estudiar las tecnologías relacionadas, y las partes de un sistema Big Data (sistemas frontales y sistemas dorsales).
- Introducirse en las características y manejo de distintas herramientas para el soporte de Big Data.
 - Diferenciar el almacenamiento masivo estructurado de los Sistemas de Gestión No-SQL
 - Introducirse en la gestión NoSQL a través de un Sistema Orientado a Documento (MongoDB). Manejo de la herramienta a nivel de manipulación de datos. Estudio práctico de las técnicas de Replicación y de Distribución de colecciones de datos para la implantación de sistemas masivamente paralelos.
 - Introducirse en otros tipos de gestión NoSQL: sistemas orientados a columna (Cassandra) y orientados a grafo (Neo4J)
 - Introducción a la suite de herramientas Hadoop

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

Bloque I: Fundamentación Teórica.

Tema 1: Introducción: Marco social y tecnológico

- Función de la Información en la sociedad de las TI
- Necesidad y tipos de sistemas de datos
- Caracterización del concepto Big Data
- Implantación de Big Data
- Aspectos legales y éticos

Tema 2: Almacenamiento y Tecnologías NoSQL

- Tecnologías de almacenamiento: estructuras y procesos
- BD transaccionales y BD analíticas
- Arquitecturas. Sistemas Distribuidos y CAP.
- Operatividad distribuida: paradigma MapReduce
- Clasificación de sistemas NoSQL

Tema 3: Integración, transformación y Limpieza

- Integración de fuentes
- Transformación y Limpieza
- Google Refine
- SPARQL

Bloque II: Herramientas para Big Data: Principales herramientas comerciales para el Almacenamiento, Informe, y Visualización

Tema 4: Back-End para BigData I: MongoDB

- Operatividad Básica en MongoDB
- Agregación en MongoDB. Pipeline y Map-Reduce.
- Replicación y Distribución en MongoDB

Tema 5: Back-End para BigData II: Neo4J

- Introducción a los Datos en Red: Grafos
- Modelos de BD en Grafo. Lenguajes.
- Property Graph DB: Neo4J

Tema 6: Back-End para BigData III: Cassandra

- Fundamentos de Cassandra
- Diseño sobre Cassandra

Tema 7: Back-End para BigData IV: Hadoop

- El ecosistema HADOOP y su instalación
- SandBox
- Funcionalidad de HADOOP
- Map-Reduce en HADOOP

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Actividades formativas:

- AF1: Clase teórica: Presentaciones teóricas acompañadas de material electrónico, como presentaciones digitales.
- AF3: Clases teórico prácticas: Combinación de clases teóricas acompañadas de la resolución de ejercicios prácticos.
- AF4: Prácticas de laboratorio: Prácticas a desarrollar en laboratorios específicos para las distintas asignaturas.
- AF5: Tutorías: Tutorías de carácter presencial y/o a distancia (videoconferencia).
- AF2: Actividades de e-learning: tutoriales, lecturas recomendadas, documentación.
- AF7: Trabajo individual del estudiante: Actividades individuales del alumno que complementan al resto de actividades (tanto presenciales como no presenciales), así como la preparación de exámenes.

Metodologías docentes:

- MD1: Exposiciones en clase del profesor con soporte de medios informáticos y audiovisuales, en las que se desarrollan los conceptos principales de la materia y se proporciona la bibliografía para complementar el aprendizaje de los alumnos.
- MD2: Lectura crítica de textos recomendados por el profesor de la asignatura: Artículos de prensa, informes, manuales y/o artículos académicos, bien para su posterior discusión en clase, bien para ampliar y consolidar los conocimientos de la asignatura.
- MD3: Resolución de casos prácticos, problemas, etc. planteados por el profesor
- MD4: Exposición y discusión en clase, bajo la moderación del profesor de temas relacionados con el contenido de la materia, así como de casos prácticos
- MD5: Elaboración de trabajos e informes de manera individual o en grupo
- MD6: Actividades específicas de e-learning, relacionadas con el carácter semi-presencial del título, actividades de auto-corrección, participación en foros, y cualquier otro mecanismo de enseñanza on-line

SISTEMA DE EVALUACIÓN

Peso porcentual del Examen Final:	60
Peso porcentual del resto de la evaluación:	40

La evaluación se llevará a cabo a través de:

- trabajo en aula (10%)

Peso porcentual del Examen Final: 60

Peso porcentual del resto de la evaluación: 40

- entregas de trabajos a lo largo del curso (30%)
- una prueba presencial de evaluación individual referida tanto a contenidos teóricos como teórico-prácticos (60%)

BIBLIOGRAFÍA BÁSICA

- Apache¿ Hadoop® <http://hadoop.apache.org/>, Apache¿ Hadoop®, 2016
- MongoDB <http://www.mongodb.org>, MongoDB, 2016

RECURSOS ELECTRÓNICOS BÁSICOS

- Google Refine . Tutorial:
- MongoDB . Documentación MongBD: <http://docs.mongodb.org/manual/>
- Sandbox . Documentación Sandbox: <http://hortonworks.com/products/hortonworks-sandbox/>