

Técnicas de Recogida de Datos (Grupo A) (MBDS)

Guía docente 2023-24

PRESENTACIÓN

Breve descripción:

• Titulación: Máster Universitario en Ciencia de Datos Masivos / Big Data Science

• Módulo/Materia: Programación y computación

• **ECTS**: 3

• Curso, semestre: curso único del Máster, primer y segundo semestre

• Carácter: obligatoria

• Profesorado: Pablo Urruchi

• Idioma: castellano

• Aula, Horario: ver cronograma

COMPETENCIAS

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:

CG2 - Explorar y explotar tanto tecnológica como estratégicamente de los datos un valor clave para diferentes empresas y organizaciones.

CG4 - Analizar los datos masivos y aportar medidas originales y novedosas para contribuir a la mejora de la competitividad de las empresas e instituciones públicas.

CG5 - Analizar los datos que se generan en el día a día, extrayendo conocimiento de los mismos, realizando predicciones y transformándolos en productos y servicios utilizando las herramientas estadísticas de Data Science.

CG6 - Trabajar en equipos de trabajo o grupos de investigación interdisciplinares de forma eficaz y colaborativa.



CG7 - Conocer y entender las herramientas habituales que se utilizan hoy día en el tratamiento de datos masivos.

PROGRAMA

Programa general de la asignatura:

- 1. Arquitectura de Computadores. Cloud Computing. Infraestructura de Cloud. OCI
- 2. SQL analítico. Oracle Autonomous Database: ADW, ATP, JSON
- 3. Big Data Cloud Products

Programa detallado de las sesiones teóricas y prácticas por días:

Sesión 1:

- 0.- Técnicas de Big Data. Indice y presentación.
- 1.- Evolución de la arquitectura de computadores, Redes de computadoras, nacimiento del Big Data. Nociones básicas (procesadores, almacenamiento, redes)
- 2.- Fundamentos de Big Data

Los temas teóricos de la primera sesión, se centran en aclarar la evolución de la arquitectura de computadoras, definición de algunos términos clave, descripción de algunos casos de uso y explicación de los entornos de realización de prácticas con VirtualBox.

La parte práctica se centra en la instalación y configuración de las máquinas virtuales que contienen algún framework para el uso de HADOOP (Hortonworks, Cloudera, Oracle etc)

Prácticas detalladas:

- 1.1- Cómo instalar VirtualBox e importar máquinas virtuales.
- 1.2- Cómo configurar el entorno de red y acceso a las máquinas virtuales.

Sesión 2:

- 3.- Ecosistema de Bigdata.
- 4.- Introducción a HADOOP y uso de HADOOP CLI
- 5.- Introducción a MAPREDUCE y YARN
- 6.- Overview de HIVE

Los temas teóricos de la segunda sesión, se centran en la explicación de las arquitecturas de computación distribuídas con HADOOP, y su capacidad de paralelización.

Prácticas detalladas:

2.1 Tutoriales sobre HortonWorks. Cargar los datos de los sensores de camiones en HDFS

Se usarán los datasets contenidos en los ficheros geolocation.csv y trucks.csv

2.2 Tutoriales sobre HortonWorks. Hive. Cómo hacer un Data ETL.

Sesión 3:



- 6.- Overview de HIVE (Continuación)
- 7.- Overview de Spark
- 8.- Nociones de Oracle NoSQL

Los temas teóricos de la tercera sesión se centran en la explicación del funcionamiento de un ETL con HIVE y en el procesamiento de datos en memoria y la paralelización con Spark. Veremos algunas nociones de la solución NoSQL de Oracle.

Prácticas detalladas:

- 3.1 Tutoriales sobre HortonWorks. Calcular el factor de riesgo con Spark.
- 3.2 Tutoriales sobre HortonWorks. Data Reporting con Zeppelin.

Sesión 4:

- 9.- Interfaces de Usuario
- 10.- Cloud Computing

Los temas teóricos de la cuarta sesión se centran en la introducción de algunos entornos de CLOUD COMPUTING como AWS, ORACLE, GOOGLE o AZURE, y en el posible uso que se puede hacer de éstos entornos.

Prácticas detalladas:

- 4.1 Configuración de cuentas en Cloud con AWS, AZURE y ORACLE
- 4.2 Configuración de máquinas y bases de datos en Cloud con ORACLE
- 4.3 Cómo acceder a los entornos de Cloud de Oracle ALWAYS FREE
- 4.4 Como crear, configurar y acceder a una base de datos Autónoma
- 4.5 Configuración de Acceso Seguro a través de herramientas Cliente a una Autónomous Database (SQLDeveloper y Data Visualization Desktop)
- 4.6 Creación de un Proyecto Completo con DV

Sesión 5:

Se abordará un caso de uso completo de (IA) Inteligencia Artificial donde se utilizará Tableau para la parte de visualización.

ACTIVIDADES FORMATIVAS

ACTIVIDADES FORMATIVAS	HORAS
Clases presenciales teóricas	10
Prácticas con ordenador	10



Trabajos dirigidos	30
Tutorías personalizadas	1
Estudio y trabajo personal	19
Examen	1

EVALUACIÓN

La nota final será un promedio con las siguientes ponderaciones.

EN LA CONVOCATORIA ORDINARIA

- Intervención en clases, seminarios y clases prácticas: 10%
- Resolución de ejercicios con el ordenador: 35%
- Trabajos individuales y/o en equipo: 20%
- Examen: 35%

EN LA CONVOCATORIA EXTRAORDINARIA

- Intervención en clases, seminarios y clases prácticas: 10%
- Resolución de ejercicios con el ordenador: 30%
- Trabajos individuales y/o en equipo: 10%
- Examen: 50%

HORARIOS DE ATENCIÓN

Contactar por correo electrónico:

• purruchi@unav.es

BIBLIOGRAFÍA

Librería de formación Online de Oracle: https://apexapps.oracle.com/pls/apex/f?p=44785:1

Documentación de BigData de Oracle: http://www.oracle.com/technetwork/database /bigdata-appliance/documentation/index.html

Página de OTN para descarga de máquinas virtuales: https://www.oracle.com/downloads/developer-vm/community-downloads.html