



UNIVERSIDAD
POLITÉCNICA
DE MADRID

PROCESO DE
COORDINACIÓN DE LAS
ENSEÑANZAS PR/CL/001



E.T.S. de Ingeniería y Sistemas
de Telecomunicación

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

595000530 - Ingeniería Big Data En La Nube

PLAN DE ESTUDIOS

59ID - Grado En Ingeniería Y Sistemas De Datos

CURSO ACADÉMICO Y SEMESTRE

2022/23 - Segundo semestre

Índice

Guía de Aprendizaje

| | |
|--|----|
| 1. Datos descriptivos..... | 1 |
| 2. Profesorado..... | 1 |
| 3. Conocimientos previos recomendados..... | 2 |
| 4. Competencias y resultados de aprendizaje..... | 2 |
| 5. Descripción de la asignatura y temario..... | 4 |
| 6. Cronograma..... | 6 |
| 7. Actividades y criterios de evaluación..... | 9 |
| 8. Recursos didácticos..... | 13 |
| 9. Otra información..... | 14 |

1. Datos descriptivos

1.1. Datos de la asignatura

| | |
|--|--|
| Nombre de la asignatura | 595000530 - Ingeniería Big Data en la Nube |
| No de créditos | 6 ECTS |
| Carácter | Obligatoria |
| Curso | Tercero curso |
| Semestre | Sexto semestre |
| Período de impartición | Febrero-Junio |
| Idioma de impartición | Castellano |
| Titulación | 59ID - Grado en Ingeniería y Sistemas de Datos |
| Centro responsable de la titulación | 59 - Escuela Técnica Superior De Ingeniería Y Sistemas De Telecomunicación |
| Curso académico | 2022-23 |

2. Profesorado

2.1. Profesorado implicado en la docencia

| Nombre | Despacho | Correo electrónico | Horario de tutorías * |
|---|-----------------|---------------------------|---------------------------------|
| Jesus Rodriguez Molina (Coordinador/a) | A4415 | jesus.rodriguez@upm.es | Sin horario. |
| Javier Martin Rueda | A4410 | javier.martin@upm.es | Sin horario. |

* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

3. Conocimientos previos recomendados

3.1. Asignaturas previas que se recomienda haber cursado

- Redes Y Servicios De Comunicaciones
- Computación En La Nube
- Programación Para Big Data

3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Conocimiento del sistema operativo UNIX a nivel usuario y a nivel administrador

4. Competencias y resultados de aprendizaje

4.1. Competencias

CB01 - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio

CB02 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio

CB03 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética

CB04 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado

CB05 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CE06 - Que los estudiantes tengan la capacidad de construir la infraestructura necesaria para la generación, transformación y transmisión de datos de cualquier fuente, volumen o velocidad.

CE07 - Que los estudiantes sepan desplegar, configurar y utilizar infraestructuras de computación conectadas de altas prestaciones para el almacenamiento y tratamiento de datos, en el ámbito de la ingeniería de telecomunicación, tanto en la nube como en sistemas locales y centros de procesado de datos.

CE08 - Que los estudiantes sean capaces de diseñar y desplegar infraestructuras virtualizadas y redes definidas por software para la comunicación, almacenamiento y tratamiento de datos.

CE09 - Que los estudiantes tengan la capacidad de aplicar las características, funcionalidades y estructura de Internet y las redes de ordenadores a la construcción de infraestructuras e integración de aplicaciones telemáticas y servicios.

CG01 - Tener capacidad de trabajar en entornos internacionales y multidisciplinares, haciendo uso de la lengua inglesa en forma oral y escrita.

CG03 - Ser capaz de explicar de forma oral o escrita las soluciones planteadas para la resolución de un problema.

CG04 - Saber identificar y utilizar las herramientas de las Tecnologías de la Información y de las Comunicaciones más adecuadas para plantear y construir soluciones a problemas

CG05 - Tener la capacidad de concebir y proponer soluciones creativas aplicando los métodos científico y de ingeniería para la definición y resolución de problemas formalizando los objetivos buscados y considerando los recursos disponibles.

CG09 - Desarrollar la capacidad de aprendizaje a lo largo de la vida (lifelong learning) para adaptarse a un sector tecnológico en continua evolución.

4.2. Resultados del aprendizaje

RA072 - Construir aplicaciones y servicios de transporte de datos (de forma distribuida, escalable y fiable), así como permitir su gestión y almacenamiento para su posterior procesamiento.

RA071 - Saber programar y analizar datos usando el paradigma de la programación funcional.

RA074 - Conocer y utilizar el modelo de programación y procesado más adecuado para cada caso concreto.

RA075 - Conocer y construir arquitecturas de procesado masivo con las tecnologías más adecuadas a cada caso tanto en la nube como en la niebla.

RA073 - Desplegar un sistema y programarlo para ser capaz de analizar datos masivos tanto en batch como en streaming.

RA076 - Utilizar herramientas para búsquedas y presentación de datos masivos.

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

Esta asignatura permitirá conocer y ser capaces de utilizar infraestructuras de Big Data para su uso y análisis. La diferencia con asignaturas anteriores es que la forma de abordar las características de big data es mediante el uso de sistemas distribuidos, lo que permite tener cualquier característica que deseemos siempre que contemos con suficientes recursos. Se analizarán diversos componentes existentes para las diversas partes del ciclo de vida de análisis de los datos.

5.2. Temario de la asignatura

1. Introducción a la ingeniería de Big Data y su programación funcional y cálculo lambda
 - 1.1. Introducción a la ingeniería de Big Data
 - 1.2. Introducción a la programación funcional (LISP)
 - 1.3. Introducción al lenguaje Scala
 - 1.4. Introducción a la programación distribuida (modelo de actores) usando Scala y Akka
2. Arquitecturas de procesados e ingeniería de datos
 - 2.1. Despliegue flexible en la nube
 - 2.2. Computación distribuida funcional: ecosistemas Hadoop y paradigma Map-Reduce
 - 2.3. Arquitecturas Spark Scala y SparkML
 - 2.4. Despliegue de servicios de alta disponibilidad : Zookeeper y protocolos de acuerdo distribuidos en la nube
3. Ciclo de vida del dato : inyección, procesado y almacenamiento
 - 3.1. Inyección masiva de datos: Paradigma Publicación-Subscripción
 - 3.2. Sistemas de procesado de datos en flujo
 - 3.3. Introducción a los lagos de datos (data lake) : NoSql y HDFS
 - 3.4. Herramientas de despliegue y orquestación de soluciones de procesado de datos
4. Despliegue en los bordes de la red y preparación para la visualización de datos
 - 4.1. Despliegue en la niebla (fog computing) en el borde de la red y 5G
 - 4.2. Herramientas de búsqueda y visualización de datos

6. Cronograma

6.1. Cronograma de la asignatura *

| Sem | Actividad presencial en aula | Actividad presencial en laboratorio | Tele-enseñanza | Actividades de evaluación |
|-----|---|--|----------------|---------------------------|
| 1 | <p>Introducción a la asignatura : metodología y evaluación Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Introducción a la Ingeniería en Big Data Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Calculo Lambda y Lisp Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> | | | |
| 2 | <p>Calculo Lambda y Lisp Duración: 03:00 LM: Actividad del tipo Lección Magistral</p> <p>Calculo Lambda y Lisp Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p> | | | |
| 3 | <p>Lenguaje Scala Duración: 03:00 LM: Actividad del tipo Lección Magistral</p> <p>Lenguaje Scala Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p> | | | |
| 4 | <p>Lenguaje Scala Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> | <p>Lenguaje Scala Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p> | | |
| 5 | <p>Actores y AKKA Duración: 03:00 LM: Actividad del tipo Lección Magistral</p> | <p>Actores y AKKA Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio</p> | | |
| 6 | <p>Despligue flexible en la nube: configuración de clusters Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Despligue flexible en la nube: configuración de clusters Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p> <p>Introducción al paradigma MAP-Reduce y el Ecosistema HADOOP Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> | | | |

| | | | | |
|----|--|---|--|--|
| 7 | Arquitectura spark Scala y Spark ML Duración: 02:00 LM: Actividad del tipo Lección Magistral | Arquitectura spark Scala y Spark ML Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio | | Examen oral de las prácticas tanto guiadas como creativas. OT: Otras técnicas evaluativas Evaluación continua Presencial Duración: 01:00 |
| 8 | Arquitectura spark Scala y Spark ML Duración: 02:00 LM: Actividad del tipo Lección Magistral | Arquitectura spark Scala y Spark ML Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio | | Examen sobre los contenidos teóricos desarrollados EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 01:00 |
| 9 | Despliegues de servicios de alta disponibilidad : Zookeeper y protocolos de acuerdo de alta disponibilidad Duración: 03:00 LM: Actividad del tipo Lección Magistral | Despliegues de servicios de alta disponibilidad : Zookeeper y protocolos de acuerdo de alta disponibilidad Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio | | |
| 10 | Inyección masiva de datos : Paradigma Publicación Suscripción Duración: 03:00 LM: Actividad del tipo Lección Magistral | Inyección masiva de datos : Paradigma Publicación Suscripción Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio | | |
| 11 | Sistemas de procesamiento de datos en flujos Duración: 03:00 LM: Actividad del tipo Lección Magistral | Despliegues de sistemas de inyección de datos Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio | | |
| 12 | Introducción al almacenamiento y lagos de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral | Introducción al almacenamiento y lagos de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio | | |
| 13 | Herramientas de búsqueda y visualización de datos Duración: 03:00 LM: Actividad del tipo Lección Magistral | Herramientas de búsqueda y visualización de datos Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio | | |
| 14 | Herramientas de búsqueda y visualización de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral | Herramientas de búsqueda y visualización de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio | | Examen oral de las prácticas tanto guiadas como creativas. OT: Otras técnicas evaluativas Evaluación continua Presencial Duración: 01:00 |
| 15 | Despliegue en la niebla (borde de la red) y 5G Duración: 02:00 LM: Actividad del tipo Lección Magistral Futuro de Big Data Duración: 02:00 LM: Actividad del tipo Lección Magistral | | | Examen sobre los contenidos teóricos desarrollados EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 01:00 |
| 16 | | | | Examen final EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final No presencial Duración: 02:00 Examen oral sobre las prácticas EP: Técnica del tipo Examen de Prácticas Evaluación sólo prueba final |

| | | | | |
|----|--|--|--|----------------------------------|
| | | | | No presencial Duración: 02:00 |
| 17 | | | | |

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación continua

| Sem. | Descripción | Modalidad | Tipo | Duración | Peso en la nota | Nota mínima | Competencias evaluadas |
|------|--|-------------------------------------|------------|----------|-----------------|-------------|--|
| 7 | Examen oral de las prácticas tanto guiadas como creativas. | OT: Otras técnicas evaluativas | Presencial | 01:00 | 25% | 0 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG09 CG04 CE09 CG01 CG03 CG05 |
| 8 | Examen sobre los contenidos teóricos desarrollados | EX: Técnica del tipo Examen Escrito | Presencial | 01:00 | 25% | 0 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 CG01 CG03 |
| 14 | Examen oral de las prácticas tanto guiadas como creativas. | OT: Otras técnicas evaluativas | Presencial | 01:00 | 25% | 0 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 |

| | | | | | | | |
|----|--|-------------------------------------|------------|-------|-----|--------|--|
| | | | | | | | CG01 CG03 |
| 15 | Examen sobre los contenidos teóricos desarrollados | EX: Técnica del tipo Examen Escrito | Presencial | 01:00 | 25% | 0 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 CG01 CG03 |

7.1.2. Evaluación sólo prueba final

| Sem | Descripción | Modalidad | Tipo | Duración | Peso en la nota | Nota mínima | Competencias evaluadas |
|-----|---------------------------------|--|---------------|----------|-----------------|-------------|--|
| 16 | Examen final | EX: Técnica del tipo Examen Escrito | No Presencial | 02:00 | 50% | 5 / 10 | CE06 CB02 CE08 CB04 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 CG01 CG03 |
| 16 | Examen oral sobre las prácticas | EP: Técnica del tipo Examen de Prácticas | No Presencial | 02:00 | 50% | 5 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 CG01 CG03 |

7.1.3. Evaluación convocatoria extraordinaria

| Descripción | Modalidad | Tipo | Duración | Peso en la nota | Nota mínima | Competencias evaluadas |
|---------------------------------|--|------------|----------|-----------------|-------------|--|
| Examen final | EX: Técnica del tipo Examen Escrito | Presencial | 02:00 | 50% | 5 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 CG01 CG03 |
| Examen oral sobre las prácticas | EP: Técnica del tipo Examen de Prácticas | Presencial | 01:00 | 50% | 5 / 10 | CE08 CB04 CE06 CB02 CB03 CB05 CB01 CE07 CG05 CG09 CG04 CE09 CG01 CG03 |

7.2. Criterios de evaluación

Los pesos relativos a cada uno de los exámenes de la evaluación progresiva han sido incluidos en los apartados anteriores. Si el alumno no supera la asignatura siguiendo la evaluación progresiva, podrá realizar una prueba de evaluación global de todo el contenido de la asignatura (el cual incluirá la parte teórica y la práctica) al finalizar el periodo de docencia. El contenido

de esta prueba será equivalente al realizado en la evaluación progresiva. No se liberan bloques de la asignatura ni en el examen global de la evaluación ordinaria ni en la extraordinaria, ni en el presente curso ni en los siguientes.

La convocatoria extraordinaria de la asignatura consistirá en: a) la evaluación mediante examen final y b) la evaluación de todas las prácticas obligatorias que no se hayan superado en la convocatoria ordinaria. Estas prácticas se deberán entregar antes de la fecha del examen extraordinario.

La asignatura se aprobará en la convocatoria extraordinaria cuando la suma ponderada de las calificaciones de la prueba final y de las prácticas sea mayor o igual a 5 puntos sobre un total de 10, y se hayan superado una nota mínima de 3 puntos sobre 10 en cada una de las partes (examen final y evaluación de las prácticas obligatorias).

Los pesos y las notas mínimas exigidas se especifican en las tablas incluidas el principio de este apartado. En caso de que un estudiante saque menos de un 3 en alguna de las partes, se le pondrá un 3 como máximo si no llega a la nota mínima, lo cual supondrá haber suspendido la asignatura.

Todos los exámenes y trabajos que se realicen deben ser fruto del trabajo personal del alumno. La copia de exámenes o trabajos prácticos supondrá aplicar a rajatabla la normativa de la UPM relativa al fraude académico. En el caso de que en el desarrollo de las pruebas de

evaluación se aprecie el incumplimiento de los deberes como estudiante universitario, el coordinador de la asignatura podrá ponerlo en conocimiento del Director del Centro, que de acuerdo con lo establecido en el artículo 74 (n) de los Estatutos de la UPM tiene competencias para "Proponer la iniciación del procedimiento

disciplinario a cualquier miembro de la Escuela o Facultad, por propia iniciativa o a instancia de la Comisión de Gobierno al Rector, en los términos previstos en los estatutos y normas de aplicación". Como se ha mencionado antes, el fraude académico está tratado en la Normativa de Evaluación Aprendizaje de la UPM, con fecha del 26 de mayo de 2022

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

| Nombre | Tipo | Observaciones |
|--|--------------|---------------|
| Common Lisp: A Gentle Introduction to Symbolic Computation | Bibliografía | |
| Scala for the Impatient | Bibliografía | |
| Functional Programming in Scala | Bibliografía | |
| Akka Concurrency | Bibliografía | |
| Hadoop: The Definitive Guide | Bibliografía | |
| Spark - The Definitive Guide: Big data processing made simple | Bibliografía | |
| Learning Spark: Lightning-fast Data Analytics | Bibliografía | |
| Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications | Bibliografía | |
| Kafka - The Definitive Guide: Real-Time Data and Stream Processing at Scale | Bibliografía | |
| Data Pipelines with Apache Airflow | Bibliografía | |
| Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine | Bibliografía | |

| | | |
|-----------------------|--------------|--|
| Moodle | Recursos web | https://moodle.upm.es |
| Infraestructura cloud | Equipamiento | Recursos de sistemas cloud públicos que permita su despliegue en un entorno real |

9. Otra información

9.1. Otra información sobre la asignatura

Esta asignatura puede contribuir a los Objetivos de Desarrollo Sostenible (ODS) 4 y 9 de Naciones Unidas, en relación con el aumento del número de personas con competencias técnicas y profesionales (meta 4.4), el desarrollo de infraestructuras fiables, sostenibles, resilientes y de calidad (meta 9.1) y ayudar a reconvertir las industrias para que sean sostenibles, utilizando los recursos con mayor eficacia, uso de tecnologías y procesos industriales limpios. Contribuir en mejorar el uso de la tecnología, en particular la tecnología de la información y las comunicaciones, para promover el empoderamiento de las mujeres (meta 5.b) y exponiendo el sesgo de sexo de algunos de los algoritmos actuales. Especialmente relevante es el objetivo 12 para minimizar la huella de carbono producida por esta tecnología.

La información contenida en esta guía es orientativa y por tanto es susceptible de modificación debido a erratas, omisiones, incidencias no previstas ocurridas durante el curso académico o si el correcto desarrollo de la asignatura así lo requiere.