# Guía Docente

## DATOS DE IDENTIFICACIÓN

| | |
|---|---|
| Titulación: | Grado en Ingeniería Matemática |
| Ámbito | Ingeniería Informática y de Sistemas. |
| Facultad/Escuela: | Escuela Politécnica Superior |
| Asignatura: | Big Data |
| Tipo: | Optativa |
| Créditos ECTS: | 3 |
| Curso: | 4 |
| Código: | 4987 |
| Periodo docente: | Octavo semestre |
| Materia: | Ciencia de Datos |
| Módulo: | Matemáticas Avanzadas y Computación |
| Tipo de enseñanza: | Presencial |
| Idioma: | Castellano |
| Total de horas de dedicación del alumno: | 75 |

| Equipo Docente | Correo Electrónico |
|---|---|
| Eusébio Daniel Rodrigues Parente | daniel.rodrigues@ufv.es |

## DESCRIPCIÓN DE LA ASIGNATURA

El concepto del Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales. Se trata del proceso de recolección de grandes cantidades de datos y su análisis para encontrar información oculta, patrones recurrentes, nuevas correlaciones, etc. El conjunto de datos con los que se trabaja es tan grande y complejo que los medios tradicionales de procesamiento son ineficaces. La asignatura ofrece una visión de los principios y necesidades logísticas de los proyectos Big Data, sobre datos con -al menos- tres "V" ya sean provenientes de fuentes públicas (fuentes abiertas) o privadas. Se explican los

sistemas de almacenamiento de datos y el flujo de conversión de raw data en datasets clasificados y categorizados, así como la infraestructura de procesamiento y almacenamiento local (on-line) y los elementos de almacenamiento y procesamiento distribuido/en nube (off-line) como Map Reduce o Hadoop, su instalación y mantenimiento.
Introduce los principios básicos de la infraestructura necesaria para afrontar proyectos Big Data y en las características principales que han dado origen al nacimiento del fenómeno Big Data, así como en el estudio de todas fases necesarias para emprender este tipo de proyectos.
A lo largo de las clases se analizarán las diversas fuentes de datos que posteriormente serán tratados, estudiando el lugar del que se obtienen y el formato de estos. Se continua con el estudio de las características y herramientas utilizadas para la ingesta de datos en los repositorios Big Data al igual que las particularidades de dichos almacenes de datos. Posteriormente se examinarán las diferentes formas en los que los sistemas Big Data procesan la información haciendo especial atención a las técnicas y herramientas disponibles para el análisis de datos. Por último, se mostrarán herramientas de visualización de datos por parte del usuario.
Durante el desarrollo de la asignatura se dará especial importancia en que el alumno comprenda que la finalidad de todo proyecto Big Data es aportar valor a las organizaciones que los implementan.

El Big Data se refiere a la necesidade procesar datos que no puede ser analizada o procesada utilizando herramientas tradicionales. Estos datos son recolectados, y pre-procesados para permitir la ejecución de otras funciones de valor añadido sobre los mismo, como analisis de patrones, correlaciones, tendencias ocultas u otra información valiosa. La cantidad de datos a tratar es tan grande y compleja que los métodos tradicionales resultan insuficientes.

En esta asignatura se proporciona una comprensión de los principios y necesidades logísticas de proyectos Big Data, incluyendo datos con al menos tres “V”, ya sean fuentes públicas o privadas. Se explican los sistemas de almacenamiento de datos, el flujo de conversión desde datos crudos hasta conjuntos clasificados y categorizados, así como la infraestructura necesaria para el procesamiento y almacenamiento local (en línea) y distribuido/en nube (fuera de línea), como Map Reduce o Hadoop, su instalación y mantenimiento.

La asignatura también introduce los principios básicos de la infraestructura necesaria para emprender proyectos Big Data y en las características principales que han dado origen al fenómeno Big Data. Además, se estudian todas las fases necesarias para llevar a cabo este tipo de proyectos.

Durante las clases se analizan diversas fuentes de datos, estudiando su lugar de origen y formato. Luego se aborda el estudio sobre las características y herramientas utilizadas para la ingesta de datos en los repositorios Big Data junto con las particularidades asociadas a dichos almacenes.

También es importante incorporar un elemento relacionado con plataformas como Hortonworks o Cloudera. Estas plataformas ofrecen soluciones integrales para gestionar entornos de big data, proporcionando herramientas para el almacenamiento, procesamiento y análisis de datos. Juegan un papel crucial en permitir a las organizaciones manejar de manera efectiva las complejidades del big data. Los estudiantes se beneficiarán al comprender cómo estas plataformas contribuyen a la implementación exitosa de proyectos de big data y el valor que aportan a las organizaciones.

Posteriormente se examinan las diferentes formas en que los sistemas Big Data procesan la información haciendo hincapié en las técnicas y herramientas disponibles para el análisis de datos. Por último, se muestran herramientas para visualizar los datos generados por el usuario.

## OBJETIVO

La materia de esta asignatura tiene como propósito situar al alumno en el entorno Big Data, con especial atención a las fases de una infraestructura Big Data.
Esta asignatura, desde un estudio teórico y práctico, permite al alumno obtener una visión global de todos los aspectos que se deben tener en cuenta en un proyecto integrado en el ecosistema Big Data, teniendo siempre presente que la finalidad de estos es aportar valor a las organizaciones que los implementan.

Los objetivos específicos que se pueden ampliar para este curso de Big Data pueden ser los siguientes:

- Comprender los conceptos y fundamentos de Big Data, incluyendo el volumen, la velocidad y el tipo de datos que se manejan, así como las tecnologías utilizadas para su gestión.
- Identificar las fases clave de una infraestructura Big Data, desde la recopilación y almacenamiento de datos hasta su análisis e interpretación.
- Conocer las herramientas y técnicas necesarias para el procesamiento y análisis de grandes conjuntos de datos (por ejemplo: Hadoop y mapreduce, Spark, NoSQL).
- Aprender cómo diseñar y desarrollar proyectos de Big Data con éxito, teniendo en cuenta las necesidades específicas de cada organización y proyecto.
- Descubrir cómo utilizar los resultados del análisis de datos para tomar decisiones informadas que agreguen valor a la organización.
  Adquirir habilidades prácticas mediante ejercicios y casos prácticos en el uso e implementación de herramientas esenciales para el manejo del Big Data.

## CONOCIMIENTOS PREVIOS

Los propios de acceso a Grado

-Conocimientos de arquitectura de ordenadores ( incluyendo paralelismo)
- Conocimientos de sistemas linux
- Conocimientos de sistemas de virtualizaciñon.
.Conocimiento de programación en lenguajes de scripting

## CONTENIDOS

- Introducción al Big Data
  En esta sección se brinda una introducción al concepto de Big Data, su importancia y beneficios. También se describen las diferentes etapas del diseño de una infraestructura para sistemas Big Data. Sistemas de virtualización / hipervisores.
- Introducción al ecosistema Hadoop+En esta sección se describirá el ecosistema de big data y se iniciará el proceso de puesta en marcha de una plataforma de prueba el ecosistema de Hadoop como Hortonworks.
- Fuentes de datos
  Se presentan las diferentes fuentes de datos en la época del Big Data. Además, se discuten los criterios para seleccionar las fuentes de datos adecuadas.
- Ingesta de datos
  En esta sección se aborda la ingesta (captura) de datos en sistemas Big Data, y se analizan las diferentes herramientas existentes para realizar dicha tarea.
- Repositorios de datos
  Se describen los diferentes sistemas disponibles para almacenar los datos capturados en un sistema Big Data.
- Procesamiento y análisis de datos
  Esta es una etapa crucial en el proceso del Big Data. Aquí se discuten las formas más comunes de procesar los datos capturados, así como las diferentes herramientas disponibles para realizar dicha tarea. Asimismo, se analizan los tipos de análisis que pueden realizarse sobre los datos y las diferentes herramientas

disponibles para llevar a cabo cada uno.

- Consumo y visualización de datos
  En esta sección se discute cómo obtener información útil a partir del conjunto masivo de datos procesados y analizados. Se detallan tanto la información histórica como la información en tiempo real; además, se analizan diversas herramientas y técnicas para visualizar los resultados obtenidos.
- Regulación y Privacidad de los datos.
  La sección de Regulación y Privacidad de los datos se enfoca en las leyes, políticas y prácticas relacionadas con la protección de la información personal y sensible. Esta sección suele ser muy importante para empresas que manejan gran cantidad de datos personales, ya que deben asegurarse de cumplir con regulaciones como el Reglamento General de Protección de Datos (RGPD) en Europa o la Ley Federal de Protección de Datos Personales en México. En general, esta sección busca garantizar que los individuos puedan tener control sobre cómo se utiliza su información personal y que las empresas sean responsables en su manejo.

## ACTIVIDADES FORMATIVAS

La metodología seguida en esta asignatura está dirigida a conseguir un aprendizaje significativo por parte del alumno de los conceptos y técnicas fundamentales de la materia.
Por ese motivo se combinan lecciones expositivas con clases prácticas y presentación de trabajos, de manera que se favorezca la participación del alumno y la interacción alumno-profesor y alumno-alumno como vía para fomentar el aprendizaje colaborativo y la capacidad de autoaprendizaje, todo ello mediante estrategias de resolución de problemas y metodologías de aprendizaje basado en proyectos. Las actividades no presenciales, que pueden ser tanto de tipo individual como colectivo, serán supervisadas por el profesor en clases y tutorías, tanto individuales como de grupo, estando encaminadas a favorecer el aprendizaje autónomo y colaborativo.
El trabajo presencial se completará con trabajo autónomo por parte del alumno, en algunos casos desarrollados en grupo, de manera que se fomente el aprendizaje cooperativo. Las actividades de carácter no presencial previstas incluyen el estudio individual, que permitirá trabajar en la fijación de los conceptos teóricos abordados en las clases expositivas correspondientes a todas las materias del módulo y adquirir la destreza práctica que se persigue con las clases prácticas, que aplicarán el aprendizaje por descubrimiento basado en problemas.
Para el desarrollo de las competencias y habilidades en esta asignatura son igualmente importantes los trabajos individuales y grupales.
Todo el estudio y trabajo realizado por el alumno será supervisado y guiado por el profesor mediante tutorías, individuales o en grupo. En algunos casos, el alumno tendrá que realizar en clase la exposición de las principales conclusiones de su estudio o trabajo, lo que permitirá el intercambio de conocimientos y experiencias entre alumnos.
Finalmente, con el fin de facilitar al alumno el acceso a los materiales y la planificación de su trabajo, así como la comunicación con el profesor y el resto de alumnos, se empleará el aula virtual, que es una plataforma de aprendizaje que ofrece diferentes recursos electrónicos para complementar, de forma muy positiva, el aprendizaje del alumno.

## DISTRIBUCIÓN DE LOS TIEMPOS DE TRABAJO

| ACTIVIDAD PRESENCIAL | TRABAJO AUTÓNOMO/ACTIVIDAD NO PRESENCIAL |
|---|---|
| 30 horas | 45 horas |
| • Clase expositiva participativas 8h<br>• Resolución de problemas o casos prácticos 6h<br>• Actividades participativas Grupales 4h | • Trabajos personal y estudio autonomo 40h<br>• Trabajo virtual en red, revisión y visionado de documentos 5h |

| | |
|---|---|
| • Seguimiento académico y actividades de evaluación 2h<br>• Practicas en laboratorio 10h | |

## RESULTADOS DE APRENDIZAJE

Conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

Utilizar las técnicas matemáticas y algorítmicas necesarias para el tratamiento de datos masivos con el fin de generar conocimiento a partir de la información que ayude en la toma de decisiones.

## RESULTADOS DE APRENDIZAJE ESPECIFICOS

Entender la relevancia de las fuentes de datos

Saber extraer y limpiar los datos necesarios para realizar el análisis

Conocer las diferentes técnicas de ingesta de datos

Entender las diferentes modalidades de procesamiento de datos.

Utilizar con soltura las herramientas de análisis utilizadas durante el curso

Trabajar con las diferentes herramientas del ecosistema Big Data.

Saber interpretar críticamente los resultados y tiene en cuenta la dimensión ética del análisis

Adquirir una visión global de toda la infraestructura Big Data

## SISTEMA DE EVALUACIÓN DEL APRENDIZAJE

El sistema de evaluación contempla cuatro tipos de pruebas:
•[1] Examen escrito teórico- práctico: presenta un peso del 35% en la nota final.
•[2] Examen escrito teórico- práctico: presenta un peso del 35% en la nota final.
•[3] Pruebas en clase, prácticas y otros trabajos relacionados con la asignatura: presenta un peso del 20% en la nota final.
•[4] Participación y presencia en clase e implicación en la asignatura: presenta un peso del 10% en la nota final.
En las tres primeras pruebas es necesario obtener un mínimo de 5 puntos sobre 10 para poder aprobar la asignatura. Aquellos alumnos que estén exentos de la obligación de asistir a clase, bien por segunda matrícula en la asignatura o sucesivas, bien por contar con autorización expresa de la Dirección del Grado, serán evaluados por el mismo tipo de pruebas. El 5% de la participación en clase podrán obtenerlo asistiendo al menos a tres tutorías con el profesor responsable de la asignatura.

Recuperación en convocatoria ordinaria: Los alumnos que no hayan alcanzado la nota mínima en el examen escrito y/o el examen de laboratorio, podrán optar a una recuperación al final del semestre. Los alumnos que no hayan alcanzado la nota mínima en las prácticas y trabajos, podrán optar a una recuperación al final del semestre.
Recuperación en convocatoria extraordinaria: Los alumnos que no hayan alcanzado la nota mínima en los exámenes, habiendo suspendido por tanto en la convocatoria ordinaria, podrán optar a una recuperación en la convocatoria extraordinaria. Los alumnos que no hayan alcanzado la nota mínima en las prácticas y trabajos, podrán optar a una recuperación en la convocatoria extraordinaria. En lo referente a la materia objeto de examen, ambas recuperaciones (ordinaria y extraordinaria), el alumno se presentará solo a las partes que tenga evaluadas por debajo de 5.

En lo referente a la materia objeto de prácticas y trabajos, en ambas recuperaciones (ordinaria y extraordinaria), el alumno se examinará de un examen práctico sobre aquellas que tenga evaluadas por debajo de 5.
La nota ponderada de la evaluación continua será un valor entre 0 y 10 y se calculará como sigue: 0,35*[1]+ 0,35*[2]+ 0,20*[3] +0,1*[4].
El alumno dispone de 6 convocatorias para superar esta asignatura. La Normativa de Evaluación de la UFV recoge todo lo relativo a los procesos de evaluación y consumo de convocatorias.
[1]Examen a mitad de cuatrimestre de carácter teórico-práctico, con cuestiones cortas, preguntas de desarrollo y ejercicios prácticos.
Este examen representará un 35% de la calificación final y evaluará la primera mitad del temario. El examen se puntuará de 0 a 10, repartiendo esta puntuación de manera equitativa entre todos los ejercicios y apartados, salvo que se indique lo contrario. Se evaluará el planteamiento de los problemas, así como la corrección, presentación e interpretación de los resultados obtenidos. [2]Examen a final de cuatrimestre de carácter teórico-práctico, con cuestiones cortas, preguntas de desarrollo y ejercicios prácticos.
Este examen representará un 35% de la calificación final y evaluará la segunda mitad del temario, si bien, debido a la relación de todos los conceptos vistos en la asignatura, se recomienda encarecidamente repasar los contenidos de la primera parte. También se puntuará de 0 a 10, repartiendo esta puntuación de manera equitativa entre todos los ejercicios y apartados, salvo que se indique lo contrario. Se evaluará el planteamiento de los problemas, así como la corrección, presentación e interpretación de los resultados obtenidos.
Ambas pruebas [1] y [2] se realizarán sin transparencias, apuntes, libros ni cualquier otro material relacionado con la asignatura. [3]Pruebas en clase, prácticas y otros trabajos relacionados con la asignatura (20% de la calificación final). Tareas individuales y en grupo de diversa índole, incluyendo prácticas y otros ejercicios o pruebas relacionados con la asignatura.
En el caso que el profesor estime oportuno, la calificación quedará afectada por la defensa oral del trabajo, al alza o a la baja, para asegurar la autoría de los trabajos. [4]Participación e implicación: 5% de la calificación final.
Se evaluarán los ejercicios y otras actividades engrupo, el interés mostrado por el alumno, concretamente se computará el índice de asistencia a tutorías tanto individuales o grupales, el grado de participación activa en las clases mediante la respuesta a preguntas del profesor, el estudio de temas avanzados no vistos en clase, la recopilación de noticias aparecidas en los medios de comunicación relacionadas con la asignatura, etc. La calificación de este apartado será un valor numérico entre 0 y 10. Aunque esta nota sea inferior a 5, no se podrá optar a recuperación. Cualquier tipo de fraude o plagio por parte del alumno en una actividad evaluable, será sancionado según se recoge en la Normativa de Convivencia de la UFV. A estos efectos, se considerará "plagio" cualquier intento de defraudar el sistema de evaluación, como copia en ejercicios, exámenes, prácticas, trabajos o cualquier otro tipo de entrega, bien de otro compañero, bien de materiales o dispositivos no autorizados, con el fin de hacer creer al profesor que son propios.

## BIBLIOGRAFÍA Y OTROS RECURSOS

### Básica

Joyanes Aguilar, Luis. Big Data: análisis de grandes volúmenes de datos en organizaciones / México :Marcombo,2014.

Caballero Roldán, Rafael. Las bases de Big Data / Madrid :Los Libros De La Catarata,2015.

### Complementaria

Mayer-Schönberger, Viktor. Big data: la revolución de los datos masivos /  Madrid :Turner,2013.