

Guía Docente: Lenguajes Avanzados de Análisis Big Data

DATOS GENERALES	
Facultad	Facultad de Ciencias y Tecnología
Titulación	Máster en Análisis Inteligente de Datos (Big Data)
Plan de estudios	2018
Materia	Tecnologías de computación en Big Data
Carácter	Obligatorio
Período de impartición	Primer Trimestre
Curso	Primero
Nivel/Ciclo	Máster
Créditos ECTS	6
Lengua en la que se imparte	Castellano
Prerrequisitos	No se prevén requisitos previos; por tanto, los requisitos serán los propios del título.

DATOS DEL PROFESORADO			
Profesor Responsable	Eugenio Marinetto Carrillo	Correo electrónico	eugenio.marinetto@ui1.es
Área		Facultad	Facultad de Ciencias y Tecnología
Perfil Profesional 2.0	LinkedIn Página personal Google Scholar		

CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DE LA ASIGNATURA**Asignaturas de la materia**

- Lenguajes Avanzados de Análisis Big Data

Contexto y sentido de la asignatura en la titulación y perfil profesional

Esta asignatura ofrece una primera visión general a dos elementos cruciales en los sistemas de Big Data: los **lenguajes de programación** y las **herramientas de procesamiento de datos**.

Por un lado, en la mayoría de sistemas de Big Data requerimos de conocimientos de lenguajes programación y algoritmia para ser capaces de especificar al sistema las tareas que queremos que realice. Dentro de los ecosistemas de Big Data y el mundo de la ciencia de datos hay distintas alternativas, pero en esta asignatura nos centraremos en **Python** por su versatilidad y uso en distintos ecosistemas de Big Data.

Por otro lado, esta asignatura también nos servirá para introducirnos a la primera herramienta para el procesamiento de grandes cantidades de datos. Nos referimos a Apache Hadoop, una herramienta de vital importancia en los ecosistemas de Big Data. Nos centraremos en conocer su **arquitectura**, su **sistema distribuido de ficheros**, y la **programación de trabajos** con Map Reduce.

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE Y RESULTADOS DE APRENDIZAJE

<p>Competencias de la asignatura</p>	<ul style="list-style-type: none"> • 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: Saber aplicar los conocimientos adquiridos y ser capaz de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio. • CB8: Ser capaz 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: Ser capaz de transmitir sus conclusiones, y los conocimientos y fundamentos 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. • CG2: Ser capaz de permanecer eficaz dentro de un medio cambiante, así como a la hora de enfrentarse con nuevas tareas, retos y personas • CG4: Ser capaz de proponer soluciones imaginativas y originales así como ser capaz de promover la innovación e identificación de alternativas contrapuestas a los métodos y enfoques tradicionales en el contexto del análisis de datos masivos o bigdata • CG5: Diseñar y desarrollar la implementación y puesta en marcha de proyectos de bigdata en diferentes áreas de aplicación social y profesional. • CG6: Ser capaz de integrarse en equipos de trabajo o investigación multidisciplinares de manera eficaz y colaborativa. • CE01: Comprender, aplicar y analizar arquitecturas y técnicas propias de bigdata para el análisis de datos estáticos y dinámicos, estructurados y no estructurados. • CE02: Identificar y utilizar herramientas software especializadas para el tratamiento de grandes volúmenes de datos en distintos contextos. • CE03: Saber diseñar y desarrollar soluciones en lenguajes y entornos de programación especializados en big data. • CE05: Diseñar, desarrollar y probar soluciones bigdata adaptadas para la captación, almacenamiento y tratamiento de grandes volúmenes de datos procedentes de diferentes contextos. • CE06: Comprender y utilizar técnicas avanzadas de visualización de datos y de experiencia de usuario para el diseño e implementación de interfaces adaptadas al usuario en los procesos de análisis de bigdata en distintos contextos.
<p>Resultados de aprendizaje de la asignatura</p>	<ul style="list-style-type: none"> • Conocer adecuadamente los principales lenguajes de programación utilizados para el análisis masivo de datos, sus sentencias y peculiaridades. • Desarrollar programas capaces de manipular bases de datos con grandes volúmenes de datos y diferentes arquitecturas. • Implementar estrategias de análisis adecuadas al tipo de datos a analizar y los resultados buscados.

PROGRAMACION DE CONTENIDOS

Breve descripción de la asignatura

En esta asignatura se verá entre otros:

- Elementos principales de Hadoop.
- Almacenamiento de datos en Hadoop.
- Gestión de recursos en Hadoop.
- Procesamiento con Map Reduce en Hadoop.
- Modelos de programación y aplicaciones para *big data*.
- Infraestructura Hadoop para Big Data

Contenidos

• **Unidad Didáctica 1: Introducción a Apache Hadoop y los lenguajes en los ecosistemas de Big Data**

1. Breve introducción al Big Data
2. Breve introducción a los sistemas distribuidos
3. Introducción a Apache Hadoop
4. Los lenguajes de programación y su relevancia en el Big Data

• **Unidad Didáctica 2: Lenguaje de programación Python básico**

1. Variables y valores
2. Estructuras condicionales
3. Estructuras iterativas y listas
4. Funciones

• **Unidad Didáctica 3: Lenguaje de programación Python avanzado**

1. Otras estructuras de datos: Tuplas y diccionarios
2. Manipulación de ficheros de texto
3. Programación orientada a objetos
4. Funciones flecha y funciones de orden superior

• **Unidad Didáctica 4: Arquitectura de Apache Hadoop y el HDFS**

1. Elementos de la arquitectura de Apache Hadoop y sus funciones
2. El sistema distribuido de ficheros HDFS
3. Distribuciones de Hadoop
4. Despliegue on-premise vs. Cloud

• **Unidad Didáctica 5: El paradigma Map Reduce**

1. Divide y vencerás y su relación la informática y los sistemas distribuidos
2. La fase de Map
3. La fase de Reduce
4. Programación de trabajos Map Reduce con Python

• **Unidad Didáctica 6: Despliegue de infraestructuras de Hadoop en AWS**

1. ¿Qué es AWS?
2. Creación de clústeres de Hadoop en AWS
3. Gestión de ficheros en AWS
4. Envío de trabajos de Map Reduce
5. Otras herramientas Big Data en AWS

METODOLOGÍA

Actividades formativas

Para cada una de las seis unidades didácticas de la asignatura, se plantearán una serie de actividades de carácter tanto formativo como sumativo con el fin de adquirir competencias y resultados de aprendizaje de la asignatura. Las actividades propuestas comprenderán:

- **Estudio de casos prácticos.** Estos se emplearán para la adquisición de competencias prácticas y que lleven al estudiante a situaciones similares a las encontradas en el mundo real. En estos casos prácticos, el estudiante aplicará los conceptos teóricos introducidos en la correspondiente unidad para la resolución de un problema. Estas actividades van encaminadas a la adquisición de resultados de aprendizaje de mayor complejidad cognitiva.
- **Contenidos teóricos.** Texto o materiales audiovisuales empleados para la introducción de nuevos conceptos en cada unidad didáctica y los aspectos más teóricos de la materia. Al final de cada unidad se adjuntarán actividades de autoevaluación que permitirán al estudiante comparar su proceso de aprendizaje actual con los resultados esperados. Además, se podrán sugerir lecturas o resolución de ejercicios teóricos para facilitar el aprendizaje del estudiante.
- **Foros de debate.** Los estudiantes debatirán sobre el impacto de determinados temas de la asignatura sobre aspectos de la vida cotidiana. Estos foros buscan fomentar el pensamiento crítico y el aprendizaje cooperativo.
- **Cuestionarios:** Cuestionario evaluables que servirán para poner a prueba los conocimientos adquiridos.

EVALUACIÓN

Sistema evaluativo

En caso de que la situación sanitaria impida la realización presencial de los exámenes con todas las garantías, la Universidad Isabel I celebrará dichas pruebas en modalidad online. Para la realización de dichos exámenes, la universidad incorporará la herramienta de proctoring a nuestra plataforma tecnopedagógica, con el objetivo de garantizar los procesos de autenticación del alumno, como el control del entorno durante el desarrollo de las pruebas de evaluación. A su vez, la Universidad Isabel I pondrá a disposición del alumnado una Unidad de Exámenes Online específica para ofrecer apoyo técnico durante todo el proceso y así solventar todas las incidencias que se puedan presentar.

El sistema de evaluación se basará en una selección de las pruebas de evaluación más adecuadas para el tipo de competencias que se trabajen. El sistema de calificaciones estará acorde con la legislación vigente (*Real Decreto 1125/2003, de 5 de septiembre, por el que se establece el sistema europeo de créditos y el sistema de calificaciones en las titulaciones universitarias de carácter oficial y de validez en todo el territorio nacional*).

El sistema de evaluación de la Universidad Isabel I queda configurado de la siguiente manera:

Sistema de evaluación convocatoria ordinaria

Opción 1. Evaluación continua

Los estudiantes que opten por esta vía de evaluación deberán realizar el **seguimiento de la evaluación continua (EC)** y podrán obtener hasta un **60 %** de la calificación final a través de las actividades que se plantean en la evaluación continua.

Además, deberán realizar un **examen final presencial (EX)** que supondrá el **40 %** restante. Esta prueba tiene una parte dedicada al control de la identidad de los estudiantes que consiste en la verificación del trabajo realizado durante la evaluación continua y otra parte en la que realizan diferentes pruebas teórico-prácticas para evaluar las competencias previstas en cada asignatura.

Para la aplicación de los porcentajes correspondientes, el estudiante debe haber obtenido una nota mínima de un 4 en cada una de las partes de las que consta el sistema de evaluación continua.

Se considerará que el estudiante supera la asignatura en la convocatoria ordinaria por el sistema de evaluación continua, siempre y cuando al aplicar los porcentajes correspondientes se alcance una calificación mínima de un 5.

Opción 2. Prueba de evaluación de competencias

Los estudiantes que opten por esta vía de evaluación deberán realizar una **prueba de evaluación de competencias (PEC)** y un **examen final presencial (EX)**.

La **PEC** se propone como una prueba que el docente plantea con el objetivo de evaluar en qué medida el estudiante adquiere las competencias definidas en su asignatura. Dicha prueba podrá ser de diversa tipología, ajustándose a las características de la asignatura y garantizando la evaluación de los resultados de aprendizaje definidos. Esta prueba supone el 50 % de la calificación final.

El **examen final presencial**, supondrá el **50 %** de la calificación final. Esta prueba tiene una parte dedicada al control de la identidad de los estudiantes que consiste en la verificación del seguimiento de las actividades formativas desarrolladas en el aula virtual y otra parte en la que realizan diferentes pruebas teórico-prácticas para evaluar las competencias previstas en cada asignatura.

Al igual que con el sistema de evaluación anterior, para la aplicación de los porcentajes correspondientes el estudiante debe haber obtenido una puntuación mínima de un 4 en cada una de las partes de las que consta la opción de prueba de evaluación de competencias.

Se considerará que el estudiante supera la asignatura en la convocatoria ordinaria por el sistema de la prueba de evaluación de competencias siempre y cuando al aplicar los porcentajes correspondientes se alcance una calificación mínima de un 5.

Sistema de evaluación convocatoria extraordinaria

Todos los estudiantes, independientemente de la opción seleccionada, que no superen las pruebas evaluativas en la convocatoria ordinaria tendrán derecho a una convocatoria extraordinaria.

La convocatoria extraordinaria completa consistirá en la realización de una **prueba de evaluación de competencias** que supondrá el **50 %** de la calificación final y un **examen final presencial** cuya calificación será el **50 %** de la calificación final.

Para la aplicación de los porcentajes correspondientes, el estudiante debe haber obtenido una nota mínima de un 4 en cada una de las partes de las que consta el sistema de evaluación de la convocatoria extraordinaria.

Los estudiantes que hayan suspendido todas las pruebas evaluativas en convocatoria ordinaria (evaluación continua o prueba de evaluación de competencias y examen final) o no se hayan presentado deberán realizar la convocatoria extraordinaria completa, como se recoge en el párrafo anterior.

En caso de que hayan alcanzado una puntuación mínima de un 4 en alguna de las pruebas evaluativas de la convocatoria ordinaria (evaluación continua o prueba de evaluación de competencias y examen final), se considerará su calificación para la convocatoria extraordinaria, debiendo el estudiante presentarse a la prueba que no haya alcanzado dicha puntuación o que no haya realizado.

En el caso de que el alumno obtenga una puntuación que oscile entre el 4 y el 4,9 en las dos partes de que se compone la convocatoria ordinaria (EC o PEC y examen), solo se considerará para la convocatoria extraordinaria la nota obtenida en la evaluación continua o prueba de evaluación de competencias ordinaria (en función del sistema de evaluación elegido), debiendo el alumno realizar el examen extraordinario para poder superar la asignatura.

Al igual que en la convocatoria ordinaria, se entenderá que el alumno ha superado la materia en convocatoria extraordinaria si, aplicando los porcentajes correspondientes, se alcanza una calificación mínima de un 5.

BIBLIOGRAFÍA Y OTROS RECURSOS

<p>Bibliografía básica</p>	<p>White, T. (2015). Hadoop: The Definitive Guide (4ª edición). O'Reilly Media.</p> <p>Es un libro clásico en la iniciación a Hadoop. Describe los conceptos fundamentales manejados en Big Data desde Hadoop. Contiene ejemplos de interés y permite ampliar contenidos tratados en esta asignatura.</p> <p>Mark Grover, Ted Malaska, Jonathan Seidman & Gwen Shapira (2015). Hadoop Application Architectures. O'Reilly Media.</p> <p>Aporta buenas prácticas a la hora de diseñar una solución Big Data de análisis de datos. Y expone casos de uso típicos sobre esta tecnología.</p>
<p>Bibliografía complementaria</p>	<p>Thilina Gunarathne, Srinath Perera (2015) Hadoop MapReduce v2 Cookbook (Second Edition). Packt (BIRMINGHAM - MUMBAI).</p> <p>Garry Turkington, Tanmay Deshpande, Sandeep Karanth (2016). Hadoop: Data Processing and Modelling. Packt (BIRMINGHAM - MUMBAI).</p>
<p>Otros recursos</p>	<ul style="list-style-type: none"> • Documentación de referencia Cloudera: https://www.cloudera.com/documentation/enterprise/latest.html • Bases de datos en AWS: https://aws.amazon.com/es/products/databases/ • SQL database en Azure: https://azure.microsoft.com/es-es/services/sql-database/ • Base de datos en Google Cloud: https://cloud.google.com/sql/?hl=es • Documentación sobre BBDD NoSQL: http://www.christof-strauch.de/nosql dbs • Introducción al lenguaje SQL https://www.w3schools.com/sql/default.asp