



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA

GUIA DOCENTE



CURSO 2019/2020

Máster en Ingeniería Informática (Plan 2018)

DATOS DE LA ASIGNATURA

Nombre:

Big Data

Denominación en inglés:

Big Data

Código:

1180406

Carácter:

Obligatorio

Horas:

Totales

Presenciales

No presenciales

Trabajo estimado:

150

60

90

Créditos:

Grupos reducidos

Grupos grandes

Aula estándar

Laboratorio

Prácticas de campo

Aula de informática

3

0

0

0

3

Departamentos:

Áreas de Conocimiento:

Tecnologías de la Información

Ciencias de la Computación e Inteligencia Artificial

Curso:

Cuatrimestre:

1º - Primero

Primer cuatrimestre

DATOS DE LOS PROFESORES

Nombre:

E-Mail:

Teléfono:

Despacho:

Márquez Hernández,
Antonio Ángel

amarquez@dti.uhu.es

959217388

ETSI 124

*Peregrín Rubio, Antonio

peregrin@uhu.es

959217653

ETSI 156

Marquez Hernandez, Francisco Alfredo	alfredo.marquez@dti.uhu.es	959217641	ETP129- Escuela Tecnica Superior Ingenieria -El Carmen
---	----------------------------	-----------	--

*Profesor coordinador de la asignatura

[Consultar los horarios de la asignatura](#)

1. Descripción de contenidos

1.1. Breve descripción (en castellano):

Versa sobre las tecnologías empleadas en esta actual óptica del tratamiento y gestión de grandes volúmenes de datos provenientes de fuentes heterogéneas como son las redes sociales, el correo electrónico, los archivos de video, etc., en volumen ilimitado, estructurados, sin estructura o estados intermedios, para proporcionar a las organizaciones técnicas de análisis que permitan adelantarse a los acontecimientos tales como las tendencias de mercado. En esta asignatura se plantean las bases de esta tecnología emergente, es decir, desde la relación con las otras tecnologías que la propician, hasta las herramientas y métodos se utilizan para implementarla en las organizaciones y sus aplicaciones, casos reales y modelos de implantación, etc.

Sus contenidos por tanto abarcan:

- Por qué surge el Big Data: Mundo de datos. Definiciones de Big Data, impacto en la sociedad actual, etc. Necesidades de nuevas arquitecturas, técnicas, algoritmos y análisis. Volumen, velocidad y variedad de datos que motiva el Big Data. Relación con Business Intelligence, Data Analytics, Minería de Datos, etc.
- Herramientas y métodos en Big Data: Paradigmas (fundamentalmente MapReduce) y sus implementaciones y frameworks: Hadoop y especialmente Spark. Ecosistemas y Bibliotecas.
- Aplicaciones, técnicas e implantaciones reales. Ejemplos típicos: telefonía, genómica, astronomía, transacciones bancarias, tráfico en internet, información web, etc.
- Aspectos legales y consideraciones éticas (legislación vigente, LOPD, privacidad, etc).

El futuro del Big Data: Áreas en evolución, tareas pendientes, hibridación con otras tecnologías, etc.

1.2. Breve descripción (en inglés):

Big Data course deals with the technologies used in the current point of view of data science: treatment and management of large volumes of data from heterogeneous sources like social networks, email, video files, etc., in unlimited volume, structured and unstructured, in order to provide the analysis that let organizations to anticipate events such as market trends, etc. In this sense, it differs from Business Intelligence, oriented to data analysis, to show how a business works in different areas and facilitate decisions, usually using data basically from the DataWarehouse structured company. In this course, it will be defined what is this emerging technology, which tools and methods are used to implement in organizations, applications, case studies and deployment models, etc.

Therefore its contents encompass:

- World of data: Why Big Data arises. Big Data definitions, impact on today's society, etc. Needs new architectures, techniques, algorithms and analysis. Volume, velocity and variety of data that motivates the Big Data.
- Tools and methods in Big Data: Platforms (Ex.: Hadoop (Open Access)), Book (eg. Mahout, Nimble, SystemML, Ricardo, Rhipe) Programming paradigms (eg.: MapReduce), algorithms, etc.
- Technical applications and actual implementations. Typical examples: telephone, genomics, astronomy, banking, Internet traffic, website information, etc. Business Intelligence and Data Analytics vs. Big Data: differences between these terms, sometimes confused and often related.
- Aspects and ethical considerations (privacy, etc.).
- The Future of Big Data: Areas evolving, hybridization with other technologies, etc.

2. Situación de la asignatura

2.1. Contexto dentro de la titulación:

Primer curso del master. Materia troncal.

2.2. Recomendaciones:

Es una materia básica de la especialidad de Big Data, por tanto, si el alumno pensase en cursar el master excepcionalmente durante más de un curso académico, esta asignatura debe escogerse antes que otras del segundo cuatrimestre.

3. Objetivos (Expresados como resultados del aprendizaje):

El alumno adquiere el conocimiento de los principios básicos y los recursos técnicos disponibles para aplicar los métodos propios del Big Data en entornos aplicados. Tras cursar esta asignatura, el alumno podrá:

- a) comprender problemas de Big Data
- b) aplicar las técnicas básicas de programación para resolver problemas de Big Data sencillos
- c) estudiar materias avanzadas en Big Data

4. Competencias a adquirir por los estudiantes

4.1. Competencias específicas:

- **CET16:** Capacidad para diseñar y evaluar sistemas operativos y servidores, y aplicaciones y sistemas basados en computación distribuida.
- **CET17:** Capacidad para comprender y poder aplicar conocimientos avanzados de computación de altas prestaciones y métodos numéricos o computacionales a problemas de ingeniería.

4.2. Competencias básicas, generales o transversales:

- **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
- **CG1:** Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la ingeniería informática
- **CG8:** Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinarios, siendo capaces de integrar estos conocimientos
- **CT1:** Gestionar adecuadamente la información adquirida expresando conocimientos avanzados y demostrando, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en el campo de estudio.
- **CT3:** Desarrollar una actitud y una aptitud de búsqueda permanente de la excelencia en el quehacer académico y en el ejercicio profesional futuro.
- **CT4:** Comprometerse con la ética y la responsabilidad social como ciudadano y como profesional, con objeto de saber actuar conforme a los principios de respeto a los derechos fundamentales y de igualdad entre hombres y mujeres y respeto y promoción de los Derechos Humanos, así como los de accesibilidad universal de las personas discapacitadas, de acuerdo con los principios de una cultura de paz, valores democráticos y sensibilización medioambiental.
- **CT5:** Utilizar de manera avanzada las tecnologías de la información y la comunicación, desarrollando, al nivel requerido, las Competencias Informáticas e Informacionales ('CI2).

5. Actividades Formativas y Metodologías Docentes

5.1. Actividades formativas:

- Sesiones de Teoría sobre los contenidos del Programa.
- Sesiones de Resolución de Problemas.
- Sesiones Prácticas en Laboratorios Especializados o en Aulas de Informática.
- Actividades Académicamente Dirigidas por el Profesorado: seminarios, conferencias, desarrollo de trabajos, debates, tutorías colectivas, actividades de evaluación y autoevaluación.

5.2. Metodologías docentes:

- Clase Magistral Participativa.
- Desarrollo de Prácticas en Laboratorios Especializados o Aulas de Informática en grupos reducidos.
- Resolución de Problemas y Ejercicios Prácticos.
- Tutorías Individuales o Colectivas. Interacción directa profesorado-estudiantes.
- Planteamiento, Realización, Tutorización y Presentación de Trabajos.
- Conferencias y Seminarios.
- Evaluaciones y Exámenes.

5.3. Desarrollo y justificación:

- En las Sesiones presenciales de Teoría se empleará en algunos casos la metodología "Clase magistral", cuando se trate de orientar y situar la asignatura en su contexto, y encuadrar las distintas líneas de la misma.
- En las Sesiones presenciales de Resolución de Problemas se propondrán, bien ejercicios, o bien completar casos de uso con ampliaciones propuestas por el profesor.
- Sesiones presenciales de Prácticas en Laboratorio de Informática, en las que el alumno adquirirá experiencia en el manejo de herramientas y en la programación de algoritmos que debe presentar y defender para su evaluación por el profesor.
- Las Actividades Académicamente Dirigidas consistirán en trabajos propuestos por el profesor para ser desarrollados por los alumnos de forma autónoma, pero con un control periódico del profesor, y la presentación final de una memoria y/o una exposición en clase por parte del alumno.
- Se programarán seminarios, cuando sea posible, y se promocionará la exposición de trabajos por parte de los alumnos para que adquieran destrezas en la presentación de los materiales que elabore.
- Las actividades no presenciales de Lectura de Contenidos consisten en la lectura propiamente de material facilitado por el profesor para este fin (a través de la plataforma Moodle) de recursos que permiten al alumno profundizar y extender su conocimiento en diferentes áreas de la materia.
- Las Tutorías Colectivas serán tutorías online a través de la plataforma Moodle en las que intervendrán tanto los alumnos, colaborativamente, como el profesor para aclarar y conducir el debate cuando sea necesario. La metodología no presencial "Tutoría en Línea" complementará, mediante sesiones de chat interactivo, las posibles necesidades de tutoría de los alumnos sin necesidad de desplazamiento al centro.
- A través de la plataforma Moodle, se pondrá a disposición de los alumnos también, cuando sea posible, de entrevistas a expertos y vídeos de sesiones magistrales de especialistas en la materia. Estas actividades no sólo son no presenciales, sino que pueden también ser comentadas en las Sesiones Teóricas presenciales en algunos casos, cuando el profesor estime que el diálogo y debate sobre las mismas pueda ser relevante.
- El Trabajo Individual Autónomo del Estudiante incluye aquellas actividades no recogidas específicamente en otras actividades, y que forman parte de las actividades que lleva a cabo no presencialmente para completar su formación, a instancias de las líneas marcadas en las Sesiones Magistrales como en las indicadas en la plataforma Moodle. Ejemplo: búsqueda documental, elaboración de esquemas, etc.
- Se podrán proponer actividades que conlleven el Trabajo Colaborativo de los estudiantes, es decir, la organización, distribución de tareas, combinación del trabajo individual o en subgrupos, todo ello orientado a la conformación final de un trabajo que precise dicha distribución entre distintos estudiantes, formentando así el trabajo en equipo para un objetivo común. En este tipo de ejercicios, cabe también la aplicación de metodologías basadas en acción, es decir, la actualización de los objetivos por los propios alumnos con el visto bueno del profesor, en función de la evolución del ejercicio llevado a cabo por el grupo de estudiantes.
- La evaluación de la asignatura se realizará, como se cita en su apartado correspondiente de esta guía, atendiendo a distintas partes de la misma y con diferente nivel de influencia en la nota final. Esto incluye, en su correspondiente porcentaje, a las sesiones presenciales de Actividades de Evaluación (Ej: exámenes) y las Actividades de Autoevaluación no presenciales (Ej: ejercicios recogidos por la plataforma Moodle puntuables).

6. Temario desarrollado:

Tema 1: Introducción a la Ciencia de los Datos y Big Data.

- Motivación del Big Data, interés de la ciencia de los datos, relación con otras materias clásicas y de actualidad. Aplicaciones.

Tema 2: Introducción al Paradigma MapReduce.

- Scale-up vs. Scale-out en HPC. Orígenes de MapReduce, en qué consiste y limitaciones.

Tema 3: Bibliotecas y Recursos para Big Data I: Apache Hadoop.

- Elementos básicos de Hadoop: sistemas de almacenamiento y procesamiento, introducción al ecosistema de Hadoop. Modelos de ejecución.

Tema 4: Bibliotecas y Recursos para Big Data II: Apache Spark.

- Fundamentos de Spark: RDDs, programación en Spark, e introducción al ecosistema de Spark.

Tema 5: Bibliotecas y Recursos para Big Data III: Apache Flink.

- Streams de datos. Introducción a Apache Flink. Arquitecturas. Aplicaciones y casos de uso.

Tema 6: Gestión de grandes volúmenes de datos.

- Elementos de los ecosistemas de Hadoop y Spark que facilitan el manejo de grandes volúmenes de datos (Hive, Pig, etc.).

Tema 7: Aprendizaje a IA en entornos de alta escalabilidad.

- Recursos para Aprendizaje automático y minería de datos en Big Data: Mahout, MLlib, KeystoneML. Recursos en la nube de Amazon (Lex, Rekognition, Machine Learning, etc.), y Google (Google Machine Learning, etc.).

Tema 8: Legislación y Big Data.

- Aspectos legales y éticos. Privacidad y LOPD en entornos de Big Data.

Tema 9: Futuro y Retos actuales del Big Data.

- Áreas en desarrollo y problemas abiertos en el ámbito de Big Data. Hibridaciones.

7. Bibliografía

7.1. Bibliografía básica:

- Big data : a revolution that will transform how we live, work, and think, Viktor Mayer-Schoenberger, Kenneth Cukier, HoughtonMifflin Harcourt, 2013 (Versión Castellano: Big Data, La revolución de los Datos Masivos)
- Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. IBM. Paul Zikopoulos, Chris Eaton. McGraw Hill Professional, 29/05/2015. <http://www-01.ibm.com/software/data/bigdata/>
- Analítica predictiva: Predecir el futuro utilizando Big Data / Eric Siegel. Madrid : Anaya Multimedia, 2013.
- Marz, N., & Warren, J. (2015). *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co.

7.2. Bibliografía complementaria:

- Del Cloud Computing al Big Data: Vision introductoria para jóvenes emprendedores. Jordi Torres i Vinnals. Editorial UOC - PID_00194204. Primera edición: septiembre 2012. Todos los derechos reservados de esta edición, FUOC, 2012. v.3.0 España de Creative Commons.
- Hadoop, Soluciones Big Data., Boris Lublinsky, Kevin Smith, Alex Yakubovich, Anaya Multimedia 2014
- Big Data, Técnicas, Herramientas y Aplicaciones, María Pérez Marque's, 2015

8. Sistemas y criterios de evaluación.

8.1. Sistemas de evaluación:

- Examen de teoría/problemas
- Defensa de Prácticas
- Defensa de Trabajos e Informes Escritos

8.2. Criterios de evaluación y calificación:

Por defecto, todos los alumnos (salvo que soliciten la Evaluación Única) serán evaluados con el sistema de Evaluación Continua en las convocatorias ordinarias (I a la III, es decir, Febrero, Septiembre y Diciembre respectivamente), el cual, es el recomendado especialmente para esta materia. No obstante, aquellos alumnos que no puedan (o no deseen) acogerse al sistema recomendado de Evaluación Continua, pueden solicitar acogerse al sistema de Evaluación Única mediante escrito firmado y entregado a través del Registro General de la Universidad (presencial o telemático) dirigido al Departamento de Tecnologías de la Información y a la atención del profesor coordinador de la materia, Antonio Peregrín Rubio (recomendándose, enviar también copia de dicho escrito al profesor por correo electrónico (peregrin@dti.uhu.es) sólo como medida complementaria, y por agilidad administrativa), donde claramente consten los datos del alumno y de la asignatura para la que solicita la Evaluación Única. La convocatoria extraordinaria para la finalización del título (Noviembre) sólo se registrará mediante el sistema de Evaluación Única, y por tanto, no se necesita realizar solicitud previa para ello.

Evaluación Continua (recomendada para las Convocatorias I a la III):

- Los conocimientos teóricos (teoría y problemas) de la materia se evaluarán mediante dos tipos de pruebas:

- 1) **un examen escrito** presencial e individual, de preguntas largas, cortas y problemas combinados, según la convocatoria del Centro para esta materia. En dicho examen no se permitirá el uso de ningún dispositivo electrónico, se le proporcionará papel y el estudiante empleará bolígrafo azul o negro propio, y no se empleará ningún elemento documental externo. El peso de este examen será de un 25%, y el alumno debe obtener al menos un 4 sobre 10 puntos para el cómputo de la nota final considerando esta parte superada.
- 2) **pruebas de evaluación mediante la plataforma de enseñanza virtual Moodle**, que consistirán en ejercicios planteados por el profesor durante el curso, generalmente de resolución breve o media, para incentivar la autonomía del estudiante en la resolución de cuestiones y problemas de la materia, para resolver de forma individual y no presencial, con plazos de entrega relativamente breves dictados en el momento de cada propuesta particular), con un peso en la evaluación final del estudiante del 20% en la nota final, requiriendo que el alumno alcance al menos una puntuación de 4 sobre 10 para que compute en la nota final considerando esta parte superada.
- Los conocimientos prácticos se evaluarán mediante **prácticas de laboratorio**, las cuales consisten en una serie de enunciados en los que se indican los objetivos a alcanzar, los medios y describen con detalle los entregables que el alumno debe enviar a la plataforma Moodle en el formato y plazo estipulados en el propio enunciado, y defendidas presencialmente en clase de prácticas de laboratorio (el alumno debe mostrar su trabajo, y responder a las cuestiones sobre el mismo que el profesor le plantee). Todas las prácticas obligatorias propuestas deben ser entregadas y defendidas presencialmente. En cada práctica, el profesor indicará si su realización es individual o por grupos (en tal caso, el tamaño y componentes del grupo debe ser supervisado y aceptado por el profesor al inicio de la práctica). No es necesario asistir presencialmente a las sesiones de prácticas que no sean las de entrega y defensa de prácticas, pero es muy recomendable hacerlo regularmente, pues se trata de una actividad diseñada para ser presencial. El peso en la evaluación de esta parte será de un 40%, y debe obtener una calificación mínima de 4 puntos sobre 10, para computar en la nota final considerando esta parte superada.
- La participación activa del estudiante en las actividades propuestas (seguimiento del estudiante), es decir, su contribución a los debates en la plataforma, foros de ideas, respuesta a ejercicios cortos, actividades académicas, etc., será también un elemento que se valorará con un peso de un 15%, a criterio de los profesores de teoría y prácticas de laboratorio. Esta parte no tiene un mínimo de puntuación establecido para considerarse superada y computar en la media ponderada de la nota final.

Los alumnos que no aprueben en la convocatoria ordinaria I (Febrero), se pueden presentar a la convocatoria ordinaria II (Septiembre) o a la convocatoria ordinaria III (Diciembre) de un mismo curso académico, y realizar la parte correspondiente a la no superada (examen escrito, pruebas de evaluación mediante la plataforma moodle, y las prácticas). Este criterio no se mantiene entre distintos cursos académicos, es decir, las partes aprobadas, no se mantienen de un curso a otro. Asimismo, la participación activa del estudiante en las actividades propuestas se mantendrá durante las tres convocatorias ordinarias del curso académico, pero no entre distintos cursos académicos.

Evaluación Única (para la Convocatoria extraordinaria para la finalización del título , y disponible para las Convocatorias I a la III):

La evaluación única consistirá en un examen presencial individual, con acreditación previa del alumno mediante DNI que deberá mostrar tantas veces se le solicite, que se celebrará en la fecha y lugar que la convocatoria del Centro fije para la asignatura. Este examen estará compuesto por una serie de partes diferentes, que se describen a continuación:

- Para valorar los conocimientos teóricos, se plantearán una serie de cuestiones (largas y cortas combinadas) escritas sobre la materia, a las que el alumno debe responder por escrito, en un tiempo máximo de 1:30 minutos. El peso de esta parte en la nota final será de un 45% (equivalente a la suma del 25% de examen escrito más un 20% de las pruebas de evaluación mediante Moodle de la modalidad de Evaluación Continua).
- Para valorar los conocimientos prácticos, se planteará la implementación de algunos algoritmos expuestos en la materia (el alumno debe hacer estas implementaciones por escrito. La duración para esta parte del examen será de 2h como máximo. El peso de esta parte en la nota final será de un 40%.
- La capacidad de tener criterio y aportar del alumno en la materia, será evaluada por el profesor oralmente, es decir, el profesor propondrá a cada alumno de forma individual (haciendo pasar a cada alumno individualmente a la sala de examen) un tema de discusión, que el alumno mantendrá con el profesor durante un tiempo máximo de 10 minutos, en los que el profesor evaluará la destreza del alumno argumentando sobre la materia. El peso en la nota final que tendrá esta parte es de un 15%.

Las partes escritas anteriormente aludidas, es decir, las cuestiones para valorar los conocimientos teóricos y prácticos, se llevarán a cabo facilitando al alumno el papel que debe rellenar empleando para ello un bolígrafo azul o negro (de su propiedad), en las condiciones ambientales más favorables que permita el aula en la que se fije la convocatoria, no permitiéndose el uso de dispositivos electrónicos (teléfonos móviles, tabletas, auriculares, ordenadores, relojes inteligentes, *wearables*, etc.). Como excepción, estarán permitidos los dispositivos electrónicos fijos tales como implantes auditivos internos, marcapasos, etc., y ante la duda, siempre se recomienda preguntar antes de llevarlos al examen al profesor. Los apuntes, libros, hojas de distinto tamaño y textos camuflados en general, no estarán permitidos puesto que no se consideran material didáctico ni documentación a utilizar admisible. La duración total del examen completo, con todas sus partes, no excederá nunca las 4 horas.

En general, tanto para la Evaluación Continua como para la Evaluación Única, se garantiza la adquisición de las competencias de la siguiente forma: mediante la evaluación (con defensa) de las prácticas de laboratorio de la Evaluación Continua, y de las preguntas del examen sobre conocimientos prácticos en el caso de la Evaluación Única, las competencias: CETI7, CETI9, CB9, CG4, CT1, CT5; mediante el examen escrito y las pruebas de evaluación mediante la plataforma de enseñanza virtual en la Evaluación Continua, y en las preguntas sobre conocimientos teóricos del examen en la Evaluación Única, las competencias: CETI7, CETI9, CB6, CB7, CG1, CG4, CT1; y mediante el seguimiento individual del estudiante en la Evaluación Continua, y en la parte oral del examen de la Evaluación Única, las competencias: CETI7, CETI9, CB6, CB7, CG4, CT4.

Nota sobre la calificación "Matrícula de Honor": si existieran más alumnos con una calificación que les permita aspirar a la matrícula de honor (10 sobre 10 puntos en la media ponderada), es decir, en caso de equidad, la calificación de Matrícula de Honor se asignará basándose en la mayor participación en clase e implicación del alumno en la asignatura (en todo tipo de sesiones), a juicio de los profesores de la asignatura.

9. Organización docente semanal orientativa:

Semanas	Grupos Grandes	Grupos Reducidos	Aula Estándar	Grupos Reducidos	Aula de Informática	Grupos Reducidos	Laboratorio	Grupos Reducidos	prácticas de campo	Pruebas y/o actividades evaluables	Contenido desarrollado
#1	1	0	1	0	0						Tema 1
#2	1	0	1	0	0						Tema 2
#3	1	0	1	0	0						Tema 2
#4	1	0	1	0	0						Tema 3
#5	1	0	1	0	0	Entrega Práctica 1					Tema 3
#6	1	0	1	0	0						Tema 4
#7	1	0	1	0	0						Tema 4
#8	1	0	1	0	0	Entrega Práctica 2					Tema 4
#9	1	0	1	0	0						Tema 5
#10	1	0	1	0	0						Tema 5
#11	1	0	1	0	0						Tema 6
#12	1	0	1	0	0	Entrega Práctica 3					Tema 6
#13	1	0	1	0	0						Tema 7
#14	1	0	1	0	0						Tema 7
#15	1	0	1	0	0	Entrega Práctica Opcional 4					Temas 8 y 9
	15	0	15	0	0						