

Máster en Ingeniería Informática (Plan 2018)

DATOS DE LA ASIGNATURA

Nombre:

Infraestructuras para Big Data

Denominación en inglés:

Big Data Infrastructure

Código:

1180414

Carácter:

Obligatorio

Horas:

	Totales	Presenciales	No presenciales
Trabajo estimado:	75	30	45

Créditos:

Grupos reducidos				
Grupos grandes	Aula estándar	Laboratorio	Prácticas de campo	Aula de informática
1.5	0	0	0	1.5

Departamentos:

Tecnologías de la Información

Áreas de Conocimiento:

Ciencias de la Computación e Inteligencia Artificial

Curso:

1º - Primero

Cuatrimestre:

Segundo cuatrimestre

DATOS DE LOS PROFESORES

Nombre:	E-Mail:	Teléfono:	Despacho:
*Peregrín Rubio, Antonio	peregrin@uhu.es	959217653	ETSI 156
Martín Moreno, Javier	javier.martin@dti.uhu.es	959217425	P132 / Escuela Técnica Superior de Ingeniería / El Carmen

*Profesor coordinador de la asignatura

Consultar los horarios de la asignatura

1. Descripción de contenidos

1.1. Breve descripción (en castellano):

Trata sobre la instalación, configuración y administración de infraestructura, tanto física (hardware real) como virtual (en la nube) para la creación de recursos computacionales para el tratamiento de datos masivo en entornos de Big Data.

Específicamente se incluye:

- Estudio de las opciones y dimensionamiento de los recursos apropiados para cada problema de Big Data dependiendo de su perfil, periodicidad de uso, etc., atendiendo al hardware necesario en cada caso, así como a las consideraciones económicas.
- Modelos de instalaciones prácticas empleando *clusters* de servidores físicamente, mediante diferentes distribuciones (Cloudera, Hortonworks, y otros) que implementan paradigmas como MapReduce (en frameworks Hadoop, Spark, etc). Herramientas de administración y monitorización del *cluster*, etc.

Utilización de las alternativas disponibles basadas en recursos de *cloud computing*: disponibilidad, contratación, administración, gestión, seguridad, legislación, etc., empleando como ejemplo algunos casos reales.

1.2. Breve descripción (en inglés):

Big Data Infrastructures Deals with instalation, configuration and management of infrastructues, phisically (real hardware) as well as virtual (in the cloud) to build computational resources for masive data treatment in Big Data enviroments.

Specifically includes:

- The study of the options and sizing of the apropiate resouces for each problem in Big Data depending on its profile, use periodicity, etc., attending to the needed hardware in each case, as well as economical considerations.
- Models of practicals instalations employing physical clusters of servers, through different distributions (Cloudera, Hortonworks, and others), implementing paradigms like MapReduce (into Hadoop, Spark, etc., frameworks). Clusters management and monitoring tools, etc.
- Use of available alternatives based on cloud computing resources: availability, contracting, administration, management, security, legislation, etc., using as an example some real cases.

2. Situación de la asignatura

2.1. Contexto dentro de la titulación:

Segundo cuatrimestre, primera mitad del mismo. Materia de especialidad Big Data.

2.2. Recomendaciones:

Haber cursado la asignatura del primer cuatrimestre Big Data.
Completar la formación de Big Data con la asignatura de Tecnicas Escalables para Análisis de Datos.

3. Objetivos (Expresados como resultados del aprendizaje):

El alumno podrá diseñar y dimensionar infraestructuras para Big Data escogiendo aprpiadamente entre sistemas hardware físicos o en la nube, y sabiendo instalarlos y ponerlos en marcha. En concreto:

- Escoger para un problema de Big Data en un entorno concreto, entre soluciones hardware o alquiladas en la nube.
- Si la opción es instalar una infraestructura hardware, saber escogerlo.
- Saber llevar a cabo la instalación y configuración del software para Big Data sobre equipos sin ninguna instalación previa.
- Alquilar servicios para Big Data en la nube.
- Administrar clusters para Big Data.

4. Competencias a adquirir por los estudiantes

4.1. Competencias específicas:

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
- **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
- **CG2:** Capacidad para la dirección de obras e instalaciones de sistemas informáticos, cumpliendo la normativa vigente y asegurando la calidad del servicio
- **CG7:** Capacidad para la puesta en marcha, dirección y gestión de procesos de fabricación de equipos informáticos, con garantía de la seguridad para las personas y bienes, la calidad final de los productos y su homologación
- **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 ('C12).

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 la 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, fomentando 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: Hardware para Big Data.

- Tipos de hardware para clusters de Big Data y elementos de diseño. Configuración de servidores: CPU, memoria RAM, dispositivos de almacenamiento y red de datos.

Tema 2: Instalaciones Software para Big Data.

- Instalación y configuración del sistema del sistema operativo, sistema de ficheros y procesamiento distribuido. Distribuciones para Big Data. Gestión del cluster.

Tema 3: Utilización de Recursos en la Nube para Big Data.

- Uso de recursos en la nube para Big Data. Migración de datos a la nube. Utilización de recursos elasticamente. Casos de uso para Google Cloud (Cloud Datastore, BigTable, etc.), Amazon AWS (EMR, Elasticsearch, Athena, Yelp, Kinesis, etc.) y Microsoft Azure (Data Lakes Analytics, HDInsight, Azure Streams, Power BI).

7. Bibliografía

7.1. Bibliografía básica:

- Cloud Analytics with Google Cloud Platform: An end-to-end guide to processing and analyzing big data using Google Cloud Platform. Sanket Thodge. ISBN: 1788839684.
- Data Science in the Google Cloud Platform. Valliappa Lakshmanan. O'Reilly. 2018.
- Expert Hadoop Administration: Managing, Tuning, and Securing Spark, YARN, and HDFS. Sam. R. Alapati. Addison-Wesley Data & Analytics. 2016.
- Mastering Azure Analytics: Architecting in the Cloud with Azure Data Lake, HDInsight, and Spark. Zoiner Tejada. O'Reilly.
- Getting Started Guide: Analyzing Big Data with AWS. Amazon Web Services.. Createspace Independent Pub. 2014. ISBN-10: 1500838969; ISBN-13:978-1500838969
- Programming Elastic MapReduce: Using AWS Services to Build an End-to-End Application, Kenvin Schmidt y Christopher Phillips,. O'Reilly, 2013. ISBN-10 : 1449363628; ISBN-13: 978-1449363628

7.2. Bibliografía complementaria:

- 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
- Del Cloud Computing al Big Data: Visión 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 Marqués, 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, Junio, 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 (Junio), 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 CB9, CG2, CG7, CT1, CT3, 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: CG1, CG2, CG7, CB7, 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: CB7, CT3, 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	2	0	2	0	0		Tema 1	
#2	2	0	2	0	0		Tema 1	
#3	2	0	2	0	0	Entrega de la Práctica 1	Tema 2	
#4	2	0	2	0	0		Tema 2	
#5	2	0	2	0	0	Entrega de la Práctica 2	Tema 3	
#6	2	0	2	0	0		Tema 3	
#7	2	0	2	0	0		Tema 3	
#8	1	0	1	0	0	Entrega Práctica 3 y Examen de Teoría	Tema 3	
#9	0	0	0	0	0			
#10	0	0	0	0	0			
#11	0	0	0	0	0			
#12	0	0	0	0	0			
#13	0	0	0	0	0			
#14	0	0	0	0	0			
#15	0	0	0	0	0			
	15	0	15	0	0			