



Guía docente

Datos Identificativos					2013/14
Asignatura (*)	Computación en Sistemas Distribuídos			Código	614473009
Titulación	Mestrado Universitario en Computación de Altas Prestacións				
Descriptorios					
Ciclo	Periodo	Curso	Tipo	Créditos	
Máster Oficial	1º cuatrimestre	Primero	Obligatoria	3	
Idioma	CastellanoGallegoInglés				
Prerrequisitos					
Departamento	Electrónica e Sistemas				
Coordinador/a	Pardo Martínez, Xoán Carlos		Correo electrónico	xoan.pardo@udc.es	
Profesorado	Pardo Martínez, Xoán Carlos		Correo electrónico	xoan.pardo@udc.es	
Web					
Descripción general	Desde hai varios anos, o uso de arquitecturas de computación paralelas foi un aspecto fundamental que permitiu o desenvolvemento de importantes áreas en múltiples campos da ciencia básica e aplicada. Con todo, o elevado custo dos sistemas paralelos tradicionais limitou o seu uso practicamente a grandes industrias e centros de investigación. Hai tempo que o uso de redes de computadores de baixo custo representa unha alternativa práctica e barata aos grandes sistemas. Como outras alternativas, as infraestruturas Grid, e máis recentemente os sistemas Cloud, aparecen como paradigmas de computación distribuída que cambian o modo no que usamos os computadores, permitindo o acceso transparente, seguro e barato a enormes recursos computacionais desde calquera lugar do mundo. O obxectivo principal desta materia é dar a coñecer estes dous novos paradigmas de computación distribuída, isto é, Grid e Cloud Computing, e instruír ao alumno na súa utilización no ámbito da computación de altas prestacións.				

Competencias de la titulación

Código	Competencias de la titulación
A5	Conocer las arquitecturas emergentes en el campo de la supercomputación.
A6	Analizar, diseñar e implementar algoritmos y aplicaciones paralelas eficientes.
A7	Adquirir conocimientos sobre las tecnologías de virtualización: instalación, configuración y utilización.
A8	Conocer las tecnologías y herramientas disponibles para la computación Grid y Cloud.
A12	Conocer las tendencias en supercomputación así como su utilización práctica en los sectores industrial, académico y público.
B1	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.
B4	Aplicar habilidades de aprendizaje que les permitan continuar estudiando de un modo que deberá de ser en gran medida autodirigido o autónomo.
B7	Capacidad de organización y planificación.
B9	Usar las nuevas tecnologías.
B10	Buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo.
C1	Expresarse correctamente, tanto de forma oral como escrita, en las lenguas oficiales de la comunidad autónoma.
C2	Dominar la expresión y la comprensión de forma oral y escrita de un idioma extranjero.
C3	Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida.
C6	Valorar críticamente el conocimiento, la tecnología y la información disponible para resolver los problemas con los que deben enfrentarse.

Resultados de aprendizaje

Competencias de materia (Resultados de aprendizaje)	Competencias de la titulación
---	-------------------------------



Habilidad para la búsqueda, selección y manejo de recursos (bibliografía, software, etc.) relacionados con la computación Grid y Cloud	AI5 AI7 AI12	BI4 BI7 BI10	CM1 CM2 CM3 CM6
Conocer y saber utilizar las herramientas básicas a nivel de usuario disponibles en entornos Grid y Cloud	AI7 AI8	BI1 BI7 BI9	
Conocer y saber utilizar alguna de las herramientas disponibles para preparar y ejecutar aplicaciones científicas en entornos Grid y Cloud	AI6 AI7 AI8	BI7 BI9	
Gestionar la seguridad en un entorno Grid	AI8	BI1 BI9	
Instalar y configurar un entorno Grid.	AI5 AI8	BI1 BI9	
Instalar y configurar un entorno de virtualización	AI5 AI7 AI8	BI1 BI9	CM3

Contenidos	
Tema	Subtema
1. Introducción a las arquitecturas distribuídas	a. Sistemas distribuídos b. Computación a gran escala: Grid y Cloud Computing
2. Computación Grid	i. Introducción * Descripción y componentes de un Grid * Ejemplos de aplicaciones e infraestructuras Grid: ESG, NeesGrid, LCG * Middleware, organizaciones y estándares Grid: OGF, WSI, OGSA ii. Middleware Grid: Globus Toolkit * Conceptos básicos sobre Globus * Conceptos básicos de seguridad en entornos Grid: PKI, certificados X.509, proxies, etc. * Componentes de Globus: seguridad, ejecución, información, etc.
3. Computación Cloud	i. Introducción * Conceptos generales * Modelos de servicio: SaaS, PaaS, IaaS * Modelos de despliegue: Público, Privado, Híbrido * Beneficios, riesgos y oportunidades * Casos de estudio ii. IaaS (Infraestructura como servicio) * Oferta de proveedores públicos * Tecnologías de virtualización * Estándares: OVF, OCCI, CDMI, ... * Middleware Open-Source para Clouds privados: OpenNebula, OpenStack, CloudStack, Eucalyptus, ... iii. Ejemplo de IaaS * Servicios básicos proporcionados * Servicios específicos: clusters virtuales, servicios para HPC,...



4. Procesamiento de grandes datos (Big Data) en la nube	<ul style="list-style-type: none"> i. Concepto de Big Data ii. Desafíos del Big Data * Almacenamiento del Big Data: bases de datos NoSQL * Procesamiento del Big Data ii. El Big Data y la nube
5. Gestión del Big Data	<ul style="list-style-type: none"> i. Introducción al modelo de programación MapReduce * Ejemplos de aplicaciones * Ejecución y optimizaciones de aplicaciones MapReduce * MapReduce en la nube ii. Implementaciones MapReduce: Hadoop * Introducción a Hadoop * Hadoop Distributed Filesystem * Aplicaciones MapReduce con Hadoop * Instalación de un cluster Hadoop iii. Ecosistema Hadoop * Introducción al ecosistema Apache Hadoop * Programación MapReduce de alto nivel: Pig, Hive * Base de datos sobre HDFS: HBase * Otras tecnologías: Zookeeper, HCatalog, Oozie, ...

Planificación			
Metodologías / pruebas	Horas presenciales	Horas no presenciales / trabajo autónomo	Horas totales
Sesión magistral	7	7	14
Trabajos tutelados	1.5	20	21.5
Prácticas de laboratorio	15	22.5	37.5
Atención personalizada	2	0	2

(*) Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos

Metodologías	
Metodologías	Descripción
Sesión magistral	En las que se expondrá el contenido teórico del temario incluyendo ejemplos ilustrativos y con el soporte de medios audiovisuales. El alumno dispondrá del material de apoyo (notas, copias de las transparencias, artículos, etc.) con anterioridad y el profesor promoverá una actitud activa, recomendando la lectura previa de los puntos del temario a tratar en cada clase, así como realizando preguntas que permitan aclarar aspectos concretos y dejando cuestiones abiertas para la reflexión del alumno.
Trabajos tutelados	Planteamiento de casos de estudio avanzados para su resolución individual por parte de los alumnos. Estos casos de estudio permitirán a los alumnos profundizar en aspectos del temario que les interesen especialmente y que no se pudieron tratar en profundidad durante las sesiones magistrales.
Prácticas de laboratorio	En las que el alumno verá el funcionamiento en la práctica de algunos de los contenidos teóricos vistos en las clases magistrales. En estas prácticas el alumno utilizará diferentes herramientas propuestas por el profesor que le permitirán profundizar y afianzar sus conocimientos sobre diferentes aspectos de la computación Grid y Cloud. Las prácticas podrían realizarse en grupos en función de la disponibilidad de los medios necesarios para su realización.

Atención personalizada	
Metodologías	Descripción



Prácticas de laboratorio	La atención personalizada durante las prácticas servirá para orientar y comprobar el trabajo que los alumnos vayan realizando según las indicaciones que se les proporcionen, dependiendo del trabajo concreto.
Trabajos tutelados	Para la realización de los trabajos tutelados los profesores proporcionarán las indicaciones iniciales necesarias, bibliografía para consulta y realizarán un seguimiento de los avances que el alumno vaya realizando para ofrecer las orientaciones pertinentes en cada caso, de modo que se asegure la calidad de los trabajos de acuerdo a los criterios que se indiquen.
Sesión magistral	Todos los profesores de la materia propondrán además un horario de tutorías en el que los alumnos podrán resolver cualquier duda relacionada con el desarrollo de la misma. Se recomienda a los alumnos la asistencia a tutorías como parte fundamental del apoyo al aprendizaje.

Evaluación		
Metodologías	Descripción	Calificación
Prácticas de laboratorio	Se realizarán prácticas individuales. Las condiciones para su realización y valoración serán las indicadas con anterioridad por el profesor y dependerán de los recursos disponibles.	60
Trabajos tutelados	El alumno realizará un trabajo individual siguiendo las indicaciones del profesor que lo orientará sobre que aspectos se tendrán en cuenta en la evaluación.	40

Observaciones evaluación
Se tendrá también en cuenta en la evaluación el seguimiento continuado de las sesiones magistrales y actividades propuestas así como la participación activa en las mismas.

Fuentes de información	
Básica	<ul style="list-style-type: none"> - Toby Velte, Anthony Velte, Robert C. Elsenpeter (2009). Cloud Computing, A Practical Approach. McGraw-Hill - Barry Wilkinson (2009). Grid Computing: Techniques and Applications. CRC Press - Tom White (2011). Hadoop: The Definitive Guide. MapReduce for the Cloud, 2º ed.. O'Reilly - Frederic Magoules, Jie Pan, Kiat-An Tan, Abhinut Kum (2009). Introduction to Grid Computing. Chapman & Hall
Complementaria	<ul style="list-style-type: none"> - Borja Sotomayor, Lisa Childers (2006). Globus Toolkit 4: Programming Java Services. Morgan Kaufmann - Anirban Chakrabarti (2007). Grid Computing Security. Springer - Chuck Lam (2011). Hadoop in Action. Manning - L. George (2011). HBase: The Definitive Guide. O'Reilly - E. Capriolo, D. Wampler, J. Rutherglen (2012). Programming Hive. O'Reilly - Alan Gates (2011). Programming Pig. O'Reilly - Jeanna Matthews, Eli M. Dow, Todd Deshane, Wenjin Hu, Jeremy Bongio, Patrick F. Wilbur, Brendan John (2008). Running Xen: A Hands-On Guide to the Art of Virtualization. Pearson Education - Ian Foster, Carl Kesselman (2004). The Grid 2: Blueprint for a new Computing Infrastructure. Morgan Kaufmann - Bernard Golden (2008). Virtualization For Dummies. Willey

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Arquitectura e Tecnoloxía de Computadores/614473001
Arquitectura e Tecnoloxía de Supercomputadores/614473002
Programación Paralela/614473004
Administración de Sistemas I/614473010
Asignaturas que se recomienda cursar simultáneamente



Asignaturas que continúan el temario

Otros comentarios

Para los fundamentos básicos de esta materia se recomienda utilizar el material recomendado por los profesores (textos, notas, artículos, etc.), con el fin de que la atención del alumno no se disperse ante la gran cantidad de recursos bibliográficos existentes en Internet sobre el tema. Con todo, para los contenidos avanzados de la materia, así como para la realización del trabajo tutelado, se recomienda la búsqueda y selección de recursos en Internet, debido a la rápida evolución de los contenidos que la materia trata (no así sus fundamentos básicos) y para disponer de una mayor riqueza en las fuentes documentales.

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías