



Teaching Guide

Identifying Data					2015/16
Subject (*)	Computación en Sistemas Distribuídos		Code	614473009	
Study programme	Mestrado Universitario en Computación de Altas Prestacións				
Descriptors					
Cycle	Period	Year	Type	Credits	
Official Master's Degree	1st four-month period	First	Obligatoria	3	
Language	SpanishEnglish				
Teaching method	Face-to-face				
Prerequisites					
Department	Electrónica e Sistemas				
Coordinador	Pardo Martínez, Xoán Carlos	E-mail	xoan.pardo@udc.es		
Lecturers	Pardo Martínez, Xoán Carlos	E-mail	xoan.pardo@udc.es		
Web					
General description	<p>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.</p>				

Study programme competences / results

Code	Study programme competences / results
A5	Coñecer as arquitecturas emerxentes no campo da supercomputación.
A6	Analizar, deseñar e implementar algoritmos e aplicacións paralelas eficientes.
A7	Adquirir coñecementos sobre as tecnoloxías de virtualización: instalación, configuración e utilización.
A8	Coñecer as tecnoloxías e ferramentas dispoñibles para a computación en sistemas distribuídos sobre unha rede e o procesamento de intentos cantidades de datos.
A12	Coñecer as tendencias en supercomputación así como a súa utilización práctica nos sectores industrial, académico e público.
B1	Poseer e comprender coñecementos que aporten unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación
B2	Que os estudantes saiban aplicar os coñecementos adquiridos e a súa capacidade de resolución de problemas en contornas novas ou pouco coñecidas dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
B5	Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun modo que haberá de ser en gran medida autodirixido ou autónomo.
B8	Capacidade de organización e planificación.
B10	Buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo.
C1	Expresarse correctamente, tanto de forma oral coma escrita, nas linguas oficiais da comunidade autónoma.
C2	Dominar a expresión e a comprensión de forma oral e escrita dun idioma estranxeiro.
C3	Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.
C6	Valorar criticamente o coñecemento, a tecnoloxía e a información dispoñible para resolver os problemas cos que deben enfrontarse.

Learning outcomes



Learning outcomes	Study programme competences / results		
Habilidade para a procura, selección e manexo de recursos (bibliografía, software, etc.) relacionados coa computación Grid e Cloud	AR5 AR7 AR12	BR1 BR5 BR8 BR10	CC1 CC2 CC3 CC6
Coñecer e saber utilizar as ferramentas básicas a nivel de usuario dispoñibles en contornas Grid e Cloud	AR6 AR7 AR8	BR8	
Coñecer e saber utilizar algunha das ferramentas dispoñibles para preparar e executar aplicacións científicas en contornas Grid e Cloud	AR7 AR8	BR2 BR8	
Xestionar a seguridade nunha contorna Grid	AR8	BR2	
Instalar e configurar unha contorna Grid.	AR5 AR8	BR2	
Instalar e configurar unha contorna de virtualización	AR5 AR7 AR8	BR2	CC3

Contents	
Topic	Sub-topic
1. Introducción ás arquitecturas distribuídas	a. Sistemas distribuídos b. Computación a grande escala: Grid e Cloud Computing
2. Computación Grid	i. Introducción * Descrición e compoñentes dun Grid * Exemplos de aplicacións e infraestruturas Grid: ESG, NeesGrid, LCG * Middleware, organizacións e estándares Grid: OGF, WSI, OGSA ii. Middleware Grid: Globus Toolkit * Conceptos básicos sobre Globus * Conceptos básicos de seguridade en contorna Grid: PKI, certificados X.509, proxies, etc. * Compoñentes de Globus: seguridade, execución, información, etc.
3. Computación Cloud	i. Introducción * Conceptos xerais * Modelos de servizo: SaaS, PaaS, IaaS * Modelos de despregue: Público, Privado, Híbrido * Beneficios, riscos e oportunidades * Casos de estudo ii. IaaS (Infraestrutura como servizo) * Oferta de provedores públicos * Tecnoloxías de virtualización * Estándares: OVF, OCCI, CDMI, ... * Middleware Open-Source para Clouds privados: OpenNebula, OpenStack, CloudStack, Eucalíptus, ... iii. Exemplo de IaaS * Servizos básicos proporcionados * Servizos específicos: clusters virtuais, servizos para HPC,...



4. Procesamento de grandes datos (Big Data) na nube	<ul style="list-style-type: none"> i. Concepto de Big Data ii. Desafíos do Big Data * Almacenamento do Big Data: bases de datos NoSQL * Procesamento do Big Data ii. O Big Data e a nube
5. Xestión do Big Data	<ul style="list-style-type: none"> i. Introducción ao modelo de programación MapReduce * Exemplos de aplicacións * Execución e optimizacións de aplicacións MapReduce * MapReduce na nube ii. Implementacións MapReduce: Hadoop * Introducción a Hadoop * Hadoop Distributed Filesystem * Aplicacións MapReduce con Hadoop * Instalación de un cluster Hadoop iii. Ecosistema Hadoop * Introducción ao ecosistema Apache Hadoop * Programación MapReduce de alto nivel: Pig, Hive * Base de datos sobre HDFS: HBase * Outras tecnoloxías: Zookeeper, HCatalog, Oozie, ...

Planning				
Methodologies / tests	Competencies / Results	Teaching hours (in-person & virtual)	Student's personal work hours	Total hours
Guest lecture / keynote speech	A5 A8 A12 B1 C1 C2 C6	7	7	14
Supervised projects	B2 B8 B10 B5 C1 C2 C6	1.5	20	21.5
Laboratory practice	A6 A7 C3	15	22.5	37.5
Personalized attention		2	0	2

(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

Methodologies	
Methodologies	Description
Guest lecture / keynote speech	Nas que se exporá o contido teórico do temario incluíndo exemplos ilustrativos e co soporte de medios audiovisuais. O alumno disporá do material de apoio (apuntamentos, copias das transparencias, artigos, etc.) con anterioridade e o profesor promoverá unha actitude activa, recomendando a lectura previa dos puntos do temario a tratar en cada clase, así como realizando preguntas que permitan aclarar aspectos concretos e deixando cuestións abertas para a reflexión do alumno.
Supervised projects	Formulación de casos de estudo avanzados para a súa resolución individual por parte dos alumnos. Estes casos de estudo permitirán aos alumnos profundar en aspectos do temario que lles interesen especialmente e que non se puideron tratar en profundidade durante as sesións maxistrais.
Laboratory practice	Nas que o alumno verá o funcionamento na práctica dalgúns dos contidos teóricos vistos nas clases maxistrais. Nestas prácticas o alumno utilizará diferentes ferramentas propostas polo profesor que lle permitirán profundar e afianzar os seus coñecementos sobre diferentes aspectos da computación Grid e Cloud. As prácticas poderían realizarse en grupos en función da dispoñibilidade dos medios necesarios para a súa realización.

Personalized attention	
Methodologies	Description



Laboratory practice Supervised projects Guest lecture / keynote speech	<p>A atención personalizada durante as prácticas servirá para orientar e comprobar o traballo que os alumnos vaian realizando segundo as indicacións que se lles proporcionen, dependendo do traballo concreto.</p> <p>Para a realización dos traballos tutelados os profesores proporcionarán as indicacións iniciais necesarias, bibliografía para consulta e realizarán un seguimento dos avances que o alumno vaia realizando para ofrecer as orientacións pertinentes en cada caso, de modo que se asegure a calidade dos traballos de acordo aos criterios que se indiquen.</p> <p>Todos os profesores da materia proporán ademais un horario de titorías no que os alumnos poderán resolver calquera dúbida relacionada co desenvolvemento da mesma. Recomendarase aos alumnos a asistencia a titorías como parte fundamental do apoio á aprendizaxe.</p>
---	--

Assessment			
Methodologies	Competencies / Results	Description	Qualification
Laboratory practice	A6 A7 C3	Realizaranse prácticas individuais. As condicións para a súa realización e valoración serán as indicadas con anterioridade polo profesor e dependerán dos recursos dispoñibles.	60
Supervised projects	B2 B8 B10 B5 C1 C2 C6	O alumno realizará un traballo individual seguindo as indicacións do profesor que o orientará sobre que aspectos se terán en conta na avaliación.	40

Assessment comments
Terase tamén en conta na avaliación o seguimento continuado das sesións maxistras e actividades propostas así como a participación activa nas mesmas.

Sources of information	
Basic	<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 (2015). Hadoop: The Definitive Guide. MapReduce for the Cloud, 4ª ed.. O'Reilly - Frederic Magoules, Jie Pan, Kiat-An Tan, Abhinit Kum (2009). Introduction to Grid Computing. Chapman & Hall
Complementary	<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 - 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 - Chuck Lam (2011). Hadoop in Action. Manning - Alan Gates (2011). Programming Pig. O'Reilly - E. Capriolo, D. Wampler, J. Rutherglen (2012). Programming Hive. O'Reilly - L. George (2011). HBase: The Definitive Guide. O'Reilly

Recommendations
Subjects that it is recommended to have taken before
Subjects that are recommended to be taken simultaneously



Subjects that continue the syllabus

Arquitectura e Tecnoloxía de Computadores/614473001
Arquitectura e Tecnoloxía de Supercomputadores/614473002
Programación Paralela/614473004
Administración de Sistemas I/614473010

Other comments

Para os fundamentos básicos desta materia recoméndase utilizar o material recomendado polos profesores (textos, apuntamentos, artigos, etc.), co fin de que a atención do alumno non se disperse ante a gran cantidade de recursos bibliográficos existentes en Internet sobre o tema. Con todo, para os contidos avanzados da materia, así como para a realización do traballo tutelado, recoméndase a procura e selección de recursos en Internet, debido á rápida evolución dos contidos que a materia trata (non así os seus fundamentos básicos) e para dispor dunha maior riqueza nas fontes documentais.

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.