UNIVERSIDADE DA CORUÑA

Teaching Guide					
Identifying Data					2019/20
Subject (*)	HPC on the Cloud			Code	614473106
Study programme	Mestrado Universitario en Computación de Altas Prestacións / High Performance Computing (Mod. Presencial)				
Descriptors					
Cycle	Period	Year		Type	Credits
Official Master's Degree	1st four-month period	First		Optional	6
Language	SpanishGalicianEnglish				
Teaching method	Face-to-face				
Prerequisites					
Department	Departamento profesorado másterEnxeñaría de Computadores				
Coordinador	Pardo Martínez, Xoán Carlos		E-mail	xoan.pardo@udc.es	
Lecturers	Fernández Pena, Anselmo Tomás Pardo Martínez, Xoán Carlos		E-mail	 xoan.pardo@udc.es	
Web	aula.cesga.es/courses/MASTERHPC7				
General description	For several years, the use of parallel computing architectures was a fundamental aspect that allowed the development of important areas in multiple fields of basic and applied science. However, the high cost of traditional parallel systems limited its use practically the large industries and research centers. The use of low-cost computer networks, as well as computing using connected infrastructures through the Internet, has been a practical and cheap alternative to large systems for some time. Thus, Cloud computing has emerged as a paradigm of distributed computing that changes the way we use computers, allowing a transparent, safe and cheap access to huge computational resources from anywhere in the world. The main objective of this subject is to show the Cloud Computing model, and how the world of High Performance Computing can use the cloud to deal with problems that, until now, were restricted to its resolution in large supercomputers. You will see different examples of how it is possible to solve problems in the field of high performance computing using distributed services and resources accessible in the cloud.				

Study programme competences	
Code	Study programme competences
A1	CE1 - Define, evaluate and select the most appropriate architecture and software to solve a problem
A6	CE6 - Know the available tools for the distributed systems computing
B2	CB7 - The students have to know how to apply the acquired knowledge and their capacity to solve problems in new or hardly explored environment inside wider contexts (or multidiscipinary) related to its area of development
B5	CB10 - The students have to possess learning skills that allows them to continue to study in a mainly self-driven or autonomous manner
B6	CG1 - Be able to search and select useful information to solve complex problems, using the bibliographic sources of the field
C1	CT1 - Use the basic technologies of the information and computing technology field required for the professional development and the long-life learning

Learning outcomes			
Learning outcomes	Study programme competences		
The student will know the basics of cloud computing and service virtualization.	AJ6		
The student will know and learn to use the basic services provided by one of the main Cloud public providers.	AJ1 AJ6		CJ1
The student will know and know how to apply the main paradigms of distributed programming used in Cloud computing.	AJ1 AJ6	BJ2	CJ1
The student will know and learn to use the services and resources available in the cloud to prepare and execute applications in the field of high performance computing.	AJ6		CJ1

The student will acquire the necessary skills for the search, selection and management of resources (bibliography, software, etc.) related to Cloud computing in the field of high performance computing.		BJ5 BJ6	

Contents	
Topic	**Sub-topic**
Introduction to Cloud Computing	
Cloud Computing services: virtual clusters	
Distributed processing models and frameworks	
Services for distributed processing in the cloud	

Planning				
Methodologies / tests	**Competencies**	**Ordinary class hours**	**Student?s personal work hours**	**Total hours**
Guest lecture / keynote speech	A1 A6	24	0	24
Laboratory practice	A1 A6 B2 B5 B6 C1	12	63	75
Supervised projects	B2 B5 B6	0	40	40
Objective test	A1 A6 B2 B6	2	0	2
Personalized attention		9	0	9
(*)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	In which the content of each topic is exposed. The student will have all the supporting material in advance (notes, slides used by the lecturer, articles, etc.). The lecturer will promote an active attitude, asking questions that will clarify specific aspects and leaving open questions for the student's reflection.
Laboratory practice	The students will resolve diverse problems which allow them to practice the topics introduced in the keynote lectures.
Supervised projects	The subject of an individual assignment will be agreed with the teacher and the student will elaborate it more deeply in an autonomous way.
Objective test	At the end of the semester there will be an exam on the contents of the subject. In this exam the topics discussed in the theoretical and practical classes will be evaluated.

Personalized attention	
Methodologies	**Description**
Supervised projects Laboratory practice	The personalized attention during the laboratory practices will serve to guide and check the students' work following to the indications they were given. To carry out the supervised assignments, students will be given the necessary initial indications and bibliographic references for consultation. During the elaboration, their progress will be monitored to offer additional guidelines to ensure the quality of the result according to predefined criteria. Every teacher will provide a tutorial schedule to resolve students' questions related to the topics of the subject. Students will be encouraged to take advantage of the tutorial sessions as a fundamental part of their learning process.

Assessment			
Methodologies	**Competencies**	**Description**	**Qualification**
Objective test	A1 A6 B2 B6	A proba poderá conter preguntas tipo test, de resposta breve ou resolución de xercicios relacionadas coa temática tratada nas sesións maxistrais e nas prácticas de laboratorio.	40

Supervised projects	B2 B5 B6	Os traballos tutelados serán sobre algún tema a convenir entre o alumno e o profesor. Valorarase o cumprimento das especificacións, a orixinalidade, a contribución personal, a metodoloxía e rigorosidade e a presentación de resultados.	20
Laboratory practice	A1 A6 B2 B5 B6 C1	Valorarase o grao de cumprimento das especificacións, a metodoloxía e rigorosidade e a presentación de resultados.	40

Assessment comments
In order to pass the subject, a minimum score of 5 out of 10 must be obtained in the practices and supervised assignment, and 5 out of 10 in the exam. Furthermore, the total subject score must be of 5 or higher. Notes of students that fail the subject are not kept for the following course. Second opportunity (July) and extraordinary The evaluation will be the same as in the first opportunity. Students will have a second deadline before the final exam to submit failed practical assignments. Condition to be considered "Absent" Do not present any assignment and do not take part in the exam. Fraud The fraud regulation of the UDC will be applied in case fraud was detected in any assignment or in the exam.

Sources of information	
Basic	- Erl T., Puttini R. and Mahmood Z. Cloud Computing, Concepts, Technology & Architecture (2013). Ed. Prentice-Hall.- White, T. Hadoop: The Definitive Guide, Storage and Analysis at Internet Scale, 4ª edición (2015). O'Reilly Media.- B. Chambers, M. Zaharia, "Spark: The Definitive Guide", O'Reilly, 2018
Complementary	- Foster, I. and Gannon, D.B. Cloud Computing for Science and Engineering (2017). The MIT Press. - Zaharia, M., Karau, H., Konwinski, A. y Patrick Wendell. Learning Spark: Lightning-Fast Big Data Analysis (2015), O'Reilly Media. - Karau, H., Warren, R,. High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark, (2017). O'Reilly Media.- Foster, I. and Gannon, D.B. Cloud Computing for Science and Engineering (2017). The MIT Press.- Zaharia, M., Karau, H., Konwinski, A. y Patrick Wendell. Learning Spark: Lightning-Fast Big Data Analysis (2015), O'Reilly Media.- Karau, H., Warren, R,. High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark, (2017). O'Reilly Media.

Recommendations
Subjects that it is recommended to have taken before
Parallel Programming/614473102
Subjects that are recommended to be taken simultaneously
High Performance Infrastructures/614473104
Subjects that continue the syllabus
Data Analytics with HPC/614473108
Other comments
Considering the strong interrelation between the theoretical and practical contents of the subject and the progressive introduction of new concepts closely related to each other, it is advisable a weekly review to make the most of the subject.An intensive use of online communication tools will be encouraged: videoconference, e-mail, chat, etc.