Course guide
270676 - CCBDA - Cloud Computing and Big Data Analytics

Last modified: 04/02/2025

Unit in charge:	Barcelona School of Informatics
Teaching unit:	701 - DAC - Department of Computer Architecture.
Degree:	MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).
	MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).

Academic year: 2024   **ECTS Credits:** 6.0   **Languages:** English

LECTURER

Coordinating lecturer:	JORDI TORRES VIÑALS - ANGEL TORIBIO GONZALEZ
Others:	Segon quadrimestre:
	ANGEL TORIBIO GONZALEZ - 11, 12

PRIOR SKILLS

General knowledge of:
- TCP/IP networking
- Operating Systems basic administration and use of the OS from the programs
- Software development

Basic knowledge of:
- Unix command line.
- Python programming language.
- Git version control system.

Warning. Students are supposed to have the above background before starting the laboratory sessions. Complimentary fast-paced materials will be provided before class to help students meet the above requirements.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEE2.1. Capability to understand models, problems and algorithms related to distributed systems, and to design and evaluate algorithms and systems that process the distribution problems and provide distributed services.
CEE2.3. Capability to understand models, problems and mathematical tools to analyze, design and evaluate computer networks and distributed systems.

Generical:
CG4. Capacity for general and technical management of research, development and innovation projects, in companies and technology centers in the field of Informatics Engineering.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.

Transversal:

CTR1. ENTREPRENEURSHIP AND INNOVATION: Capacity for knowing and understanding a business organization and the science that rules its activity, capability to understand the labour rules and the relationships between planning, industrial and commercial strategies, quality and profit. Capacity for developping creativity, entrepreneurship and innovation trend.

CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.

CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Basic:

CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.

TEACHING METHODOLOGY

Lectures, reading and discussion of technical and research papers, Presentation of topics (and papers) by students. Laboratory sessions and a practical class project.

Students are required to bring their laptop to carry out the laboratory sessions and practical class project.

This subject is taught only in English language.

LEARNING OBJECTIVES OF THE SUBJECT

1.Present the student with new execution environments required to manage the computing resources and simplify the development and integration of the different types of applications and services at nowadays Internet-scale systems.
2.Collaborate in the design, implementation and presentation of a cloud computing environment that is required for a class project.
3.Find and understand useful information to create innovative solutions.

STUDY LOAD

Type	Hours	Percentage
Hours large group	24,0	16.00
Guided activities	6,0	4.00
Self study	96,0	64.00
Hours medium group	12,0	8.00
Hours small group	12,0	8.00

Total learning time: 150 h

CONTENTS

Lectures: Cloud Computing fundamentals

Description:
Fundamental concepts: The effect of scale on system properties.
---- Issues in large-scale systems: virtualization, service orientation and composition, availability, locality, performance and adaptation.
---- Models for large-scale systems: system models for analysis, architectural models and service/deployment models.
---- Scaling techniques: basic techniques, scalable computing techniques for architectural models.
---- Middleware and Applications: computing, storage, web, content distribution, Internet-scale systems or services.
---- Environment and applications requirements.

Laboratory sessions: Practical view of Cloud Computing

Description:
Big Data Analytics in the Cloud
---- APIs: The Doors in the Cloud
---- Current required layers in Big Data Software Stack
---- New Software requirements for Advanced Analytics
---- New Hardware requirements for Advanced Analytics

Assigment: Experimental part

Description:
Development of a prototype application using Cloud service offerings (such as AWS, Google AppEngine, Open Stack, OpenNebula)
---- Development of a prototype application using advanced analytics services either provided regarding APIs or Software as a Service.

ACTIVITIES

Introduction

Description:
Cloud Computing Definition. Service Oriented Architectures. Web Services.Business considerations
Business considerations

Full-or-part-time: 6h
Self study: 4h
Theory classes: 2h

Cloud Computing Architecture

Description:
Technology. Architecture. Modelling and Design.

Specific objectives:
1

Related competencies :
CEE2.1. Capability to understand models, problems and algorithms related to distributed systems, and to design and evaluate algorithms and systems that process the distribution problems and provide distributed services.
CEE2.3. Capability to understand models, problems and mathematical tools to analyze, design and evaluate computer networks and distributed systems.

Full-or-part-time: 6h
Self study: 4h
Theory classes: 2h

Virtualization

Description:
Foundations. Grid, cloud and virtualization.

Full-or-part-time: 6h
Self study: 4h
Theory classes: 2h

Data Storage

Specific objectives:
1

Related competencies :
CEE2.1. Capability to understand models, problems and algorithms related to distributed systems, and to design and evaluate algorithms and systems that process the distribution problems and provide distributed services.
CEE2.3. Capability to understand models, problems and mathematical tools to analyze, design and evaluate computer networks and distributed systems.

Full-or-part-time: 12h
Self study: 8h
Theory classes: 4h

Cloud Services

Full-or-part-time: 12h
Self study: 8h
Theory classes: 4h

Cloud Security

Full-or-part-time: 6h
Self study: 4h
Theory classes: 2h

Service Oriented Architectures

Full-or-part-time: 6h
Self study: 4h
Theory classes: 2h

Cloud Tools

Full-or-part-time: 12h
Self study: 8h
Theory classes: 4h

Cloud Applications

Full-or-part-time: 9h
Self study: 6h
Theory classes: 3h

Future Trends

Full-or-part-time: 4h
Self study: 2h
Theory classes: 2h

Collaborative class project

Specific objectives:
2, 3

Related competencies :
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CEE2.1. Capability to understand models, problems and algorithms related to distributed systems, and to design and evaluate algorithms and systems that process the distribution problems and provide distributed services.
CG4. Capacity for general and technical management of research, development and innovation projects, in companies and technology centers in the field of Informatics Engineering.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR1. ENTREPRENEURSHIP AND INNOVATION: Capacity for knowing and understanding a business organization and the science that rules its activity, capability to understand the labour rules and the relationships between planning, industrial and commercial strategies, quality and profit. Capacity for developping creativity, entrepreneurship and innovation trend.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.

Full-or-part-time: 34h
Self study: 30h
Guided activities: 4h

Lab: Basic knowledge toolbox

Full-or-part-time: 2h
Guided activities: 1h 30m
Laboratory classes: 0h 30m

Lab: Doors in the cloud

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Lab: Content Delivery Network

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Lab: Extract and analyze data

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Lab: Interact with users and services

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Lab: Monitoring and Security

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Lab: Data storage

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Lab: Web Services

Full-or-part-time: 2h
Guided activities: 0h 30m
Laboratory classes: 1h 30m

Discuss: Virtualization

Full-or-part-time: 4h
Self study: 2h
Guided activities: 1h
Laboratory classes: 1h

Discuss: Cloud providers comparison

Full-or-part-time: 4h
Self study: 2h
Guided activities: 1h
Laboratory classes: 1h

Discuss: Federated Cloud Computing

Full-or-part-time: 4h
Self study: 2h
Guided activities: 1h
Laboratory classes: 1h

Discuss: Cloud governance

Full-or-part-time: 4h
Self study: 2h
Guided activities: 1h
Laboratory classes: 1h

Discuss: Future trends

Full-or-part-time: 4h 42m
Self study: 2h
Guided activities: 1h 30m
Laboratory classes: 1h 12m

GRADING SYSTEM

Students will be evaluated on their participation in class, laboratory sessions, class attendance, reading and presenting reports and papers and assignments on specific topics.

The final grade for the course is the weighted average of the grades for the following components obtained en each part of the course:
· Lab sessions: 30%
· Papers Reading/Presentation and homework: 20%
· Course Projects: 30%
· Final exam: 20%

In order to be able to publicly defend the course project, students must have attended at least 70% of the classes and teams must have delivered on time the activities that have been planned during the course. The course project is the result of teamwork, which will be reflected in the grade given to the group as a whole. Each member of the group will be responsible for part of the project and might be graded individually on his or her contribution.

BIBLIOGRAPHY

Basic:
- Murugesan, S.; Bojanova, I. Encyclopedia of cloud computing. Chichester: Wiley, 2015. ISBN 9781118821978.
- Srinivasan, A.; Suresh, J. Cloud computing: a practical approach for learning and implementation. Uttar Pradesh: Pearson India Education Services Pvt. Ltd, 2014. ISBN 9788131776513.