COURSE DATA

Data Subject	
Code	36435
Name	Data storage infrastructure
Cycle	Grade
ECTS Credits	6.0
Academic year	2020 - 2021

Study (s)

Degree	Center	Acad. year	Period
1406 - Degree in Data Science	School of Engineering	2	First term

Subject-matter

Degree	Subject-matter	Character
1406 - Degree in Data Science	12 - Computer Science	Obligatory

Coordination

Name	Department
BOLUDA GRAU, JOSE ANTONIO	240 - Computer Science
PEREZ CONDE, CARLOS	240 - Computer Science
PEREZ SOLANO, JUAN JOSE	240 - Computer Science

SUMMARY

Data Storage Infrastructure is taught as a component of Computing in the first semester of the second year of the Degree in Data Science.

The course is based on programming knowledge previously acquired on Programming Fundamentals and Data Structures and Algorithms and serves as the basis for later courses such as Networks and Security, Parallel Programming and Big Data.

Students learn about operating systems, how these systems rely on the computer’s physical components, what services they offer, and how they are administered. Specifically, the various levels of local storage are studied, beginning with storage devices and ending with local file systems. Also addressed are virtualization techniques for resources and the hardware used in Big Data applications.

The theory classes will be taught in Spanish. The language for the practical and laboratory classes will be specified in the course guidelines available on the website for this degree.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

Students are recommended to have passed the courses on Programming Fundamentals and Data Structures and Algorithms.

OUTCOMES

1406 - Degree in Data Science

- Students must have acquired knowledge and understanding in a specific field of study, on the basis of general secondary education and at a level that includes mainly knowledge drawn from advanced textbooks, but also some cutting-edge knowledge in their field of study.
- Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.
- (CG01) Knowledge of basic subjects and technologies that enable students to learn new methods and technologies, and to provide them with versatility to adapt to new situations.
- (CG04) Ability to work in a multidisciplinary group in a multilingual environment and to communicate, orally and in writing, knowledge, procedures, results and ideas related to data science.
- (CT03) Ability to defend your own work with rigor and arguments and to expose it in an adequate and accurate way with the use of the necessary means.
- (CT05) Ability to evaluate the advantages and disadvantages of different methodological and / or technological alternatives in different fields of application.
- (CE04) To know and use the different models of data storage and database management systems using programming languages for the definition, query and handling of data.
- (CE08) Ability to understand, select and use the infrastructure and the techniques used to handle mass data, according to criteria of efficiency, scalability, security, error tolerance and adaptation to the production environment.

LEARNING OUTCOMES

RA1 - Know and be able to use the services and tools offered by operating systems. (CB1, CB4, CG1, CG4, CT3, CT5, CE4, CE8)

RA2 - Know and be able to use the various levels of local storage, from physical storage devices to file systems. (CB1, CB4, CG1, CG4, CT3, CT5, CE4, CE8)

RA3 - Know the usual techniques for resource virtualization and be able to select and use them. (CB1, CB4, CG1, CG4, CT3, CT5, CE4, CE8)

RA4 - Know the basic structure of the hardware used in Big Data. (CB1, CG1, CE4, CE8)

DESCRIPTION OF CONTENTS

1. Introduction to operating systems

Concept of operating system
Processes and threads
Memory management
I/O Management
File systems
Linux OS administration

2. Local storage management

Physical devices
Disk scheduling
RAID
Logical volumes
Block cache
Advanced file systems

3. Virtualization

Virtualization types and techniques
Virtual machines
Containers
Operating systems for virtualization

4. Hardware for Big Data

Servers for data processing
Specialized storage systems
Specialized backup systems

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	30,00	100
Laboratory practices	20,00	100
Classroom practices	10,00	100
Development of group work	5,00	0
Development of individual work	5,00	0
Study and independent work	20,00	0
Readings supplementary material	10,00	0
Preparation of evaluation activities	20,00	0
Preparing lectures	20,00	0
Preparation of practical classes and problem	10,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

The various teaching activities for this course are:

- Theoretical activities. The topics taught in the theory classes will provide a global and integrating vision, analyze the key and most complex aspects in detail, and encourage student participation (competences CB1, CB4, CG1, CT5, CE4, CE8).
- Practical activities. These activities will complement the theoretical activities. The basic concepts will be applied and expanded with the knowledge and experience students acquire by completing their assignments. These activities include solutions to problems and questions discussed in the classroom, discussion sessions, problem solving and other exercises previously worked on by the students, laboratory practice, oral presentations, conferences and scheduled individual or group tutoring sessions (competences CB1, CB4, CG1, CG4, CT3, CT5, CE4, CE8).
- Individual work. This includes completion (outside the classroom) of monographs, literature research, questions, problems, and studying for classes and exams. These are done individually and are intended to promote autonomous learning (competences CB1, CG1, CT3, CT5, CE4, CE8).

- Small-group work. Group work completed outside the classroom in small groups of 2-4 students comprising assignments, questions and problems. This type of activity complements individual activities and fosters teamwork (competences CB1, CB4, CG4, CT3, CT5, CE4, CE8).

EVALUATION

FIRST EXAMINATION SITTING

The course is evaluated in two ways. In the first scenario, both the in-class quizzes and the final quiz are used to calculate the final grade. In the second scenario, the in-class quizzes do not count. The final grade will be the higher of the grades obtained using these two scenarios.

At the first examination sitting, the final grade will comprise the following:

(TP) Evaluation of theory and problems.

This component will account for 75% of the final grade. Students will need to obtain a minimum of 4.5 points out of 10 to be able to pass the course. Students who do not obtain this minimum score will fail the course. The score for this component will depend on the student's participation and three quizzes taken during the course, as follows:

- (SE3) Continuous Assessment (CA) is based on the student's participation and involvement in the teaching and learning process. Both attendance and in-class activities are taken into account. This component cannot be re-taken. (Competences CB1, CB4, CG1, CG4, CT3, CT5, CE4, CE8).
- (SE1) Quizzes. Two in-class quizzes will be conducted, one in the first half of the semester (called T1) and one in the second half of the semester (called T2). One final quiz will be conducted outside class times during the exam period (called T3). (Competences CB1, CG1, CT5, CE4, CE8).

Each quiz will address all the course content taught up to the date of that quiz.

In order not to penalize students who perform better in the final quiz than in the in-class quizzes, TP is calculated as follows:

TP = Maximum(0.2 * CE + 0.1 * T1 + 0.25 * T2 + 0.45 * T3, 0.2 *CE + 0.8 * T3)

(L) (SE2) Evaluation of the laboratory activities (L). This depends on the student's fulfilment of the objectives of the laboratory sessions (competences CB1, CB4, CG1, CG4, CT3, CT5, CE4, CE8).

Laboratory work is conducted in pairs. The grade for this work accounts for 25% of the final grade. As with TP, students will need to obtain a minimum of 4.5 points out of 10 to pass the course. Students who do not obtain this minimum score will fail the course. All laboratory sessions will carry the same weight in the final grade.

Students who are unable to attend any laboratory session must submit their lab work to the lab instructor before the lab session is held. Delivery should be in person during tutorial hours. Students should be prepared to answer questions on their work and to re-do parts of it (with minor changes) in real-time. Work delivered in this way will be penalized by subtracting 20% from the grade obtained.

The algorithm used to compute the final grade is as follows:

If TP <4.5 or L <4.5: final_grade = Minimum (TP, L)
In any other case: final_grade = 0.75 * TP + 0.25 * L

SECOND EXAMINATION SITTING

At the second examination sitting, a delivery period for submitting laboratory work will be opened. Students should submit laboratory work in person during tutorial hours. Students should be prepared to answer questions on their work and to re-do parts of it (with minor changes) in real-time. Work delivered in this way will be penalized by subtracting 30% from the grade obtained.

A final examination (FE) will be also be held. This exam will replace quiz T3.

Except for these two differences, the module will be evaluated in the same way as at the first examination sitting (the CA mark will be the same as for the first examination sitting).

To apply for an advanced examination sitting, students must have previously taken the course and obtained the minimum mark required to evaluate the practical laboratory activities (L). This procedure is intended to reconcile a student’s right to an advanced examination sitting with the teaching methodology and evaluation criteria for this course.

In all cases the evaluation system will be governed by the University of Valencia’s regulations on grading and assessment for bachelor’s degrees and master’s degrees, which is available at:

https://webges.uv.es/uvTaeWeb/MuestraInformacionEdictoPublicoFrontAction.do?accion=inicio&idEdictoSeleccionado=5639

REFERENCES

Basic

- "Operating System Concepts". Abraham Silberschatz, Peter Baer Galvin y Greg Gagne. John Wiley & Sons.

- "Operating Systems" [Recurs electrònic]. William Stallings. Pearson Education
- "Unix and Linux System Administration Handbook". Evi Nemeth, Garth Snyder, Trent R. Hein, Ben Whaley. Prentice Hall.

Additional

- "Sistemas Operativos". William Stallings. Prentice Hall.
- "Sistemas Operativos. Una visión aplicada". Jesús Carretero, Félix García, Pedro de Miguel y F. Pérez. McGraw-Hill.
- "Fundamentos de Sistemas Operativos". Abraham Silberschatz, Peter Baer Galvin y Greg Gagne. John Wiley & Sons.

ADDENDUM COVID-19

This addendum will only be activated if the health situation requires so and with the prior agreement of the Governing Council

The teaching methodology of this course will follow the Teaching Model approved by the Academic Committee for the Degree in Data Science (https://go.uv.es/cienciadatos/ModelDocentGCD). In the event of a closure of the School of Engineering that totally or partially affects the lessons, classes will be replaced by virtual sessions following the established schedule. If the closure affects any face-to-face examination task, it will be replaced by a virtual exam of a similar nature using the tools supported by the University of Valencia. Percentages of all examination tasks will be maintained, as established in this guide.