COURSE DATA

Data Subject	
Code	34695
Name	Next Generation Information Systems
Cycle	Grade
ECTS Credits	6.0
Academic year	2020 - 2021

Study (s)

Degree	Center	Acad. year	Period
1400 - Degree in Computer Engineering	School of Engineering	4	Second term

Subject-matter

Degree	Subject-matter	Character
1400 - Degree in Computer Engineering	16 - Optional subject	Optional

Coordination

Name	Department
FUERTES SEDER, ARIADNA	240 - Computer Science
VES CUENCA, ESTHER DE	240 - Computer Science

SUMMARY

The volume of data stored today is growing exponentially. In 2000 the stored information was around 800,000 petabytes (PB). It is estimated that it will reach 35 zetabytes (ZB) in 2020. Only Twitter generates more than 7 terabytes (TB) of data every day, Facebook 10 TB and some companies generate terabytes of data every hour of every day of every year. It is no longer uncommon to find companies with storage capacities of the order of petabyte data. The drivers of this growth are mainly the development of the Web, the growing number of social network providers, the development of cloud services and general model of "software as a service".

In parallel, the ability to treat and explore all this information overload and apply the techniques of converting raw data into meaningful information are critical in any organization, company or institution today. Therefore, the demand for highly scalable systems has also grown exponentially to accommodate the volume of data and processing needs.

Although relational databases have demonstrated the ability to successfully adapt to the growing storage needs, now more than ever we need new paradigms of persistence solid, reliable and scalable able to meet the demands of the new information systems. There are two strategies that are being taken to resolve this issue:

1. 1. On the one hand, traditional relational systems can be tuned to meet the needs of scalability through parallelization (higher speed), distribution (higher storage capacity and process) and federation and replication (horizontal scalability by simply node addition).
2. 2. On the other hand, the introduction of new paradigms such as NoSQL and Big Data, which try to attack the problem by proposing new storage structures more versatile, even at the expense of the loss of some functionality traditionally considered good.

This course will study and analyze the advantages and disadvantages of both strategies and present the new persistence technologies that the IT professional must know.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

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

Other requirements

It is recommended to have attended Databases in the 2nd year and Database Management Systems 3rd year.

OUTCOMES

1400 - Degree in Computer Engineering

- SI1 - Ability to integrate ICT solutions into business processes in order to meet the information needs of organisations, thus enabling them to achieve their goals effectively and efficiently and providing them with competitive advantage.
- SI3 - Ability to actively participate in the specification, design, implementation and maintenance of information and communication systems.

LEARNING OUTCOMES

- · Understanding the motivation of management systems parallel databases, distributed and federated.
- · Having the basic fundamentals to analyze different architectures, implications on performance, speed, scalability and security of different partitioning strategies and evaluate and optimize queries over distributed data.
- · Understand what is data warehousing and why is it important for decision support.
- · Understanding the multidimensional data model and the type of data that facilitates analysis.
- · Understand what is data mining and have the basics of the techniques used in data mining.

- · Knowing the MapReduce algorithm and its use in the rapid analysis of vast amounts of data.
- · Know the different mass storage systems known as NoSQL systems and have the basics to define, design and implement information systems using these systems.

DESCRIPTION OF CONTENTS

1. Advanced database management systems

Distributed Databases.
Parallel databases.
Federated databases.
Replication.

2. Decision support systems

Data warehouses
Data mining

3. Big data and NoSQL

Terms.
NoSQL Databases.
MapReduce.
Distributed Storage.

4. Information Retrieval and XML Data

Information Retrieval.
XML Data Model.
XQuery.
Efficient evaluation of XML queries.

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	20,00	0
Study and independent work	10,00	0
Readings supplementary material	5,00	0
Preparation of evaluation activities	15,00	0
Preparing lectures	10,00	0
Preparation of practical classes and problem	25,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

The learning process will be based on a combination of lectures, problem resolution sessions and practical activities (computer based laboratories). It will be complimented with the student personal work.

- · Lectures will have a duration of 90 minutes, and different subjects will be taught, trying to promote the students participation.
- · Practical activities will be based on problems resolution in the classroom. The duration of the sessions will be 90 minutes. Some of the planned activities are:
  - o Problems resolution sessions
  - o Seminars in regular classrooms and computer facilities
  - o Debates, problems resolution and exercises previously worked by students
  - o Individual tutorials
- · Laboratory sessions will focus on the resolution of problems introduced in the regular lectures, with a duration of 150 minutes.
- · Individual personal work for promoting the autonomous learning process, based on these aspects:
  - o Preparation of lectures and reading of recommended texts
  - o Problems resolution
  - o Homework to be evaluated by the lecturer
  - o Laboratory sessions preparation with anticipation
- · Teamwork. Development of activities in small groups, inside and outside the classroom.

The virtual learning platform of the UVEG (Aula Virtual) will be used as a support to the teaching process and the student-lecturer communication. All the course material will be made available in this platform.

EVALUATION

The work that the students are expected to do can be classified into 2 types:

1. Autonomous self-learning work.

2. Supervised work.

1. Autonomous work of self-learning consists of the activities that the student performs outside the teaching time, independently or directed by the teacher, and is oriented to acquire knowledge, skills and skills in a self-taught way.

Mainly, they will be activities that the teacher indicates should be done during the course (recommended reading, problem solving, investigation of certain subjects, etc.) but that will not be evaluated directly (qualified) by the teacher, although they will have an impact on other activities that Yes they will be evaluated and qualified by the teacher, as they are the case of the debates and presentations realized during the hours of teaching or the conducting of examinations.

2. The supervised work consists of all the activities that the student will carry out at the request of the teacher and that the teacher will monitor to evaluate the degree of improvement of the subject. These works will be of 3 types:

A) Individual problems or activities

B) In group

C) Computing classroom practice, compulsory attendance.

The characteristics of these works are:

• Must be evaluated by the teacher.

• They must have a delivery date, and / or be made in person at the time of being raised, which would require assistance in such cases.

• They are mandatory for the student to apply continuous assessment.

At first call, the student will carry out during the course several individual objective tests during the teaching hours and that will consist of theoretical-practical questions as well as problems or debates (N_Test) that will not be recoverable.

In addition, different activities corresponding to the activities developed in the laboratory (N_Activities) will have to be delivered.There will finally be a test (N_Examen).

The final qualification, in first call, will be obtained by means of the following weighting of the different activities and tests:

Final Note = 20% N_Examen + 40% N_Prueba + 40% N_Activities

In second call the final qualification will be obtained based on the formula:

Final Score = 60% N_Examen + 40% N_Activities

It will be necessary to obtain at least 5 out of 10 in each of the parts in order to mediate the note.

In any case, the evaluation of this subject will be done in compliance with the University Regulations in this regard, approved by the Governing Council on 30th May 2017 (ACGUV 108/2017).

This assessment starts from the premise that teaching at the University of Valencia is, by definition, on-campus lecture delivery method. In this sense, the student should be aware that attendance at both the theoretical and practical lectures is essential for proper monitoring of the contents of the course. The student must also consider the possibility to enroll part time (except in the case of students who register for the first time), when it is unable to attend all courses (60 credits). However, there is an exception for those students that justify it and request it. They have the possibility of being assessed without attending to all or part of the lectures. For these cases, students should proceed as follows:

· At the beginning of the course, student should inform to lecturer responsible for the course, the incidence that makes her/him unable to attend the class. This must be adequately justified in documentary form.

· The lectures in charge, in the light of this information, will decide the possibility of evaluation without full or partial assistance to the lectures.

Students who are in this situation must submit for evaluation all work required by the lecturer (not necessarily the same to those required for the course) and may also be called to defend them orally to the lecturer, and conduct a knowledge test. The weight of the final grade work will be 50% and the test the remaining 50% knowledge.

REFERENCES

Basic

- Ramez A. Elmasri, Shmkant B. Navathe Fundamentos de Sistemas de Bases de Datos. Addison Wesley, 2002.
- R. Ramakrishnan, J. Gehrke Database Management Systems. McGraw-Hill, 2000.
- S. Tiwari. Proffesional NoSQL. Wrox, 2011.

- T. White. Hadoop: the definitive guide. Wrox, 2009

Additional

- Referencia c1: Priscilla Walmsley. XQuery. OReilly Media, Inc., 2007
- H. Inmon. Building the Data Warehouse, John Wiley & Sons, 2005
- D. Miner. MapReduce design patterns. O'Reilly 2012
- Christof Strauch. NoSQL databases. http://www.christof-strauch.de/nosqldbs.pdf
- C. Adamson. Mastering data warehouse aggregates. Willey, 2006

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 for this subject will follow the model approved by the Academic Committee of the GII / GIM degrees (https://links.uv.es/catinfmult/modeloDocent). If the facilities are closed because of COVID-19 pandemics, the scheduled lectures will be replaced by synchronous online sessions within the assigned time slots of the course, using the tools provided by the university.

If the facilities need to be closed due to the pandemics causing any of the evaluation exercises to be held at ETSE-UV, these exercises will be substituted by equivalent exercieses held online using the tools provided by the university. The weights for each activity will remain the same as specified in the teaching guide.