BASES DE DADOS AVANÇADAS

Master in Engenharia Informática

Code: 28501

Main Scientific Area: Information Systems and Artificial Intelligence

Lecturer: Joaquim Gonçalves Pereira da Silva

Language of Instruction: Portuguese

Regime: S1

Contact Hours: 60h Total Workload: 100h

ECTS: 6,0

Objectives

The curricular unit of Advanced Databases intends to consolidate the knowledge of relational databases and introduce students to other databases paradigms. The curricular unit will include classes about data analysis and processing, based on big data and cloud platforms, discussing operational and analytical aspects.

Learning Outcomes

At the end of the course unit, the students should be able to:
1. Develop efficient and secure data handling solutions based on relational database management systems;
2. Understand the models and database systems oriented to documents, key/value, columns, and graphs;
3. Model and implement solutions on database management systems suited to the type of data and functional requirements;
4. Develop efficient solutions for the processing of large volumes of data for operational and analytical purposes.

Course Contents

1. Relational databases
- Database Management systems
- Indexes and partitioning
- Triggers, stored procedures, and functions
- Query optimization and monitoring
2. NoSQL and Big Data
- Big Data Data and file formats for
- NoSQL databases: concepts and practical examples
- Map-Reduce and HDFS
- Hive Data Warehouse: structure, manipulation, and practical examples
3. Big Data Analytics
- Loading data to Hive
- Extraction, transformation, and data analysis with Spark

Recommended Bibliography

Kleppmann, M. (2020). Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. O'Reilly Media, Inc. ISBN 978-1449373320
Luu, H. (2018). Beginning apache spark 2: with resilient distributed datasets, spark SQL, structured streaming and spark machine learning library. Apress. ISBN 978-1484235805
Sadalage, P. J., Fowler, M. (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley. ISBN 978-0321826626

Sharda, R., Delen, D., Turban, E. (2017). Business Intelligence, Analytics, and Data Science: A Managerial Perspective. 4th edition, Pearson. ISBN 978-0134633282
Shaw, S., Vermeulen, A. F., Gupta, A., Kjerrumgaard, D. (2016). Practical Hive: A Guide to Hadoop's Data Warehouse System. Apress. ISBN 978-1484202722
Silberschatz, A., Korth, H. F., Sudarshan, S. (2019). Database System Concepts, 7th Ed. McGraw-Hill Education. ISBN 978-0078022159

Learning and Teaching Methods

The program covers the scope of all curricular unit objectives, complementing the theoretical contents with a practical component of implementation and testing. The realization of exercises in classes and the development of team projects throughout the lessons will consolidate the theoretical knowledge and provide students with technical skills highly valued in the job market.

Assessment Methods

There will be theoretical and practical classes. In the first, will be presented concepts, approaches, methodologies, and fundamental technologies. In the practical classes, problems will be solved with proposed solutions, promoting a participatory discussion of the students. The evaluation is composed of the development of two applied projects.

The Final Grade (FG) is the heavy average calculated according to the following expression:

FG = Project 01 Grading (P01) * 50%. + Project 02 Grading (P02) * 50%

The evaluation includes a session of individual defence of the projects, being the classification given individually to each member of the working group. To get approval in the curricular unit, students need to achieve a minimum score of 9.0 (on a scale of 0 to 20) in each project (P01 >= 9 P02 >=9). The assessment in the exam period consists of the submission of a new project followed by its presentation and defense.