ULPGC Universidad de Las Palmas de Gran Canaria

GUÍA DOCENTE *CURSO: 2025/26*

40386 - BIG DATA

CENTRO: *180 - Escuela de Ingeniería Informática*
TITULACIÓN: *4045 - Grado en Ciencia e Ingeniería de Datos*
ASIGNATURA: *40386 - BIG DATA*
CÓDIGO UNESCO: **TIPO:** *Obligatoria* **CURSO:** *3* **SEMESTRE:** *1º semestre*
CRÉDITOS ECTS: *6* **Especificar créditos de cada lengua:** **ESPAÑOL:** *0* **INGLÉS:** *6*

Enlace a la MEMORIA DE VERIFICACIÓN

https://www2.ulpgc.es/plan-estudio/4045/40/verificacion/4

REQUISITOS PREVIOS RESPECTO A ASIGNATURAS DE LA TITULACIÓN

Desarrollo de Aplicaciones para Ciencia de Datos
Ingeniería del Software
Fundamentos de programación

CONTENIDOS TEÓRICOS, PRÁCTICOS Y DE LABORATORIO

1. Theoretical concepts of Big Data
1.1 Complexity Management
1.2 Monitoring and Performance Engineering
1.3 Architectures for Big Data
10 horas
Kleppmann, Martin. Designing Data-Intensive Applications: The Big Ideas behind Reliable, Scalable, and Maintainable Systems. O'Reilly, 2017

2. Distributed and parallel Programming.
2.1 Parallel Programming.
2.2 Map Reduce.
2.3 Distributed file systems.
2.4 Development of applications for execution on Big Data clusters
2.5 Vector Programming.
40 horas
Kleppmann, Martin. Designing Data-Intensive Applications: The Big Ideas behind Reliable, Scalable, and Maintainable Systems. O'Reilly, 2017

3. Applications
3.1 Databases
3.2 Online data analysis
3.3 Machine learning
3.4 Graph analysis
10 horas
Kleppmann, Martin. Designing Data-Intensive Applications: The Big Ideas behind Reliable, Scalable, and Maintainable Systems. O'Reilly, 2017

Practical Content:
The practical content is directly related to the theoretical content, such that practical exercises based on the theory covered in each block will be carried out on a weekly basis.

The 2030 Agenda establishes that, in order to achieve sustainable development, action must be taken against poverty in all its forms and dimensions, inequality must be addressed, efforts must be made to preserve the planet, promote a sustainable economy, and foster social inclusion. Therefore, the commitment to sustainability must systematically address the economic, social, and environmental dimensions. The Sustainable Development Goals (SDGs) clearly present education as a key instrument for advancing sustainability.

This course adheres to the curricular sustainability guidelines issued by CRUE and ULPGC itself through the 2030 Agenda and its 17 SDGs. The competencies and content of “Big Data” will incorporate topics and references related to SDGs 5, 9, and 12; and both the teaching methodology and assessment will be guided, wherever possible, by best practices in sustainability. In particular, the syllabus and activities carried out in the course will be approached with a focus on those SDGs most closely related to Big Data.

In reviewing and correcting this section, the professor made use of artificial intelligence.

EVALUACIÓN:

Criterios y sistemas de evaluación

This course's assessment will be based on the following elements:

FE1. Class Participation and Attendance
Class participation is gauged by the student's contribution to the class through asking questions, making comments, and solving exercises proposed by the professor. Attendance is also monitored, which may include the student answering questions related to that day's class. These checks may be conducted in writing or through the virtual campus platform.

FE2. Class Exercises
This element of assessment includes the exercises submitted by students during class and any additional assignments. These contribute to the continuous assessment grade.

FE3. Work Assignment
Each student is required to complete several individual assignments throughout the course, in addition to participating in a group project that will be formally presented and defended. The evaluation of the individual assignments and the group project takes into account not only the quality and accuracy of the results obtained, but also the clarity and thoroughness of the written reports, as well as each student’s ability to explain and defend their contributions. This comprehensive evaluation provides a deeper understanding of each student’s grasp of the subject matter and their ability to apply their knowledge effectively, both independently and collaboratively.

Each of these components (FE1, FE2, and FE3) will be used to calculate the student's final grade, with different weights applied depending on whether the student follows the continuous evaluation model.
We aim to implement a continuous evaluation model, which is predicated on regular class attendance. This model focuses not only on final outcomes but also on progress throughout the course, enabling real-time feedback and iterative learning adaptations. Regular attendance is vital

for this model as it facilitates active, self-directed learning and allows for consistent formative assessment. This enables us to address learning gaps in a timely, individualized manner, thereby promoting more effective and personalized learning experiences.

In order to be included in the final grade calculation, all assignments must achieve a minimum score of 5 out of 10.

The final grade for the subject will be expressed numerically, in accordance with the provisions of Article 5 of Royal Decree 1125/2003, of September 5 (BOE September 18), which establishes the European Credit System and the Grading System in official university degrees valid throughout the national territory.

Grading System:
0.0 - 4.9 Fail
5.0 - 6.9 Pass
7.0 - 8.9 Notable
9.0 - 10 Sobresaliente

If the student has used AI in any of their activities, they must explicitly indicate it within them.

Criterios de calificación

The evaluation system applies to all convocatories throughout the academic term.

When a student participates in the continuous evaluation model (and attends classes regularly), the final grade will be calculated using the following formula:
Final Grade = 0.1FE1 + 0.5FE2 + 0.4*FE3

In contrast, when a student does not participate in the continuous evaluation model, the final grade will be computed as follows:
Final Grade = 0.6FE2 + 0.4FE3

In the event that a student does not submit the required assignment (FE3), or fails to complete the exercises or attend the exam (FE2), they will be assigned a 'No presentado' grade.

PLANIFICACIÓN SEMANAL

Week 1: Theoretical concepts of Big Data
Theory: The course and its first topic, complexity management, are introduced. 2 hours
Practice: Students are introduced to the big data oriented group project and start the initial phase of implementation. 2 hours
Work: 6 hours

Week 2: Theoretical concepts of Big Data
Theory: Complexity management discussions continue, techniques for managing it are introduced. 2 hours
Practice: Students learn to use Git for effective version control in their big data project and continue the implementation phase. 2 hours
Work: 6 hours

Week 3: Theoretical concepts of Big Data
Theory: Solutions applying complexity management techniques are presented. 2 hours
Practice: Students learn Git Flow methodology to streamline their group project work, while

continuing development. 2 hours
Work: 6 hours

Week 4: Theoretical concepts of Big Data
Theory: The second topic, performance engineering, is introduced. 2 hours
Practice: Students learn to use Maven for building and managing the big data project. 2 hours
Work: 6 hours

Week 5: Theoretical concepts of Big Data
Theory: Discussion on performance engineering issues continues, and performance engineering techniques are introduced. 2 hours
Practice: Students learn JMH for Java performance tuning in their big data project. The first iteration of the project is presented, and the next iteration is discussed. 2 hours
Work: 6 hours

Week 6: Distributed and parallel Programming
Theory: Solutions applying performance engineering techniques are presented. 2 hours
Practice: Students learn Docker for creating and managing containers in their big data project, and continue development. 2 hours
Work: 6 hours

Week 7: Distributed and parallel Programming
Theory: The third topic, parallel programming, is introduced. 2 hours
Practice: Students learn Java threading mechanisms for optimizing their big data project and continue the development phase. 2 hours
Work: 6 hours

Week 8: Distributed and parallel Programming
Theory: The discussion continues on performance issues in single-threaded applications, and parallel programming is introduced. 2 hours
Practice: Students learn Hazelcast for in-memory data grid capabilities in their big data project. 2 hours
Work: 6 hours

Week 9: Distributed and parallel Programming
Theory: Solutions applying parallel programming techniques are presented. 2 hours
Practice: Students learn Docker Compose to manage multi-container Docker applications in their big data project. 2 hours
Work: 6 hours

Week 10: Distributed and parallel Programming
Theory: The fourth topic, distributed programming (MapReduce), is introduced. 2 hours
Practice: Students learn MapReduce in Python for processing and generating big data sets in their project. The second iteration of the project is presented, and the next iteration is discussed. 2 hours
Work: 6 hours

Week 11: Distributed and parallel Programming
Theory: The discussion continues on performance issues in non-distributed applications and introduces distributed programming using MapReduce. 2 hours
Practice: Students learn MapReduce in Java with Hadoop for the processing of large data sets in their big data project. 2 hours
Work: 6 hours

Week 12: Distributed and parallel Programming
Theory: Solutions applying distributed programming techniques with MapReduce are presented. 2 hours
Practice: Students learn Nginx for load balancing in their big data project, and continue the development phase. 2 hours
Work: 6 hours

Week 13: Distributed and parallel Programming
Theory: The fifth topic, vector programming, is introduced. 2 hours
Practice: Students learn Google Cloud for deploying, scaling, and diagnosing production issues in their big data project. 2 hours
Work: 6 hours

Week 14: Distributed and parallel Programming
Theory: The discussion continues on performance issues in the experiments and vector programming using CUDA is introduced. 2 hours
Practice: Students learn to use Google Cloud with Hadoop MapReduce and distributed file usage in their big data project, and continue development. 2 hours
Work: 6 hours

Week 15: Distributed and parallel Programming
Theory: Solutions applying vector programming techniques are presented. 2 hours
Practice: Students present the third iteration of their big data oriented group project. 2 hours
Work: 6 hours

PROFESORADO

Dr./Dra. María Dolores Afonso Suárez *(COORDINADOR)*
Departamento: *260 - INFORMÁTICA Y SISTEMAS*
Ámbito: *075 - Ciencia De La Comp. E Intel. Artificial*
Área: *075 - Ciencia De La Comp. E Intel. Artificial*
Despacho: *INFORMÁTICA Y SISTEMAS*
Teléfono: *928458727* **Correo Electrónico:** *marilola.afonso@ulpgc.es*

Dr./Dra. José Juan Hernández Cabrera *(RESPONSABLE DE PRACTICAS)*
Departamento: *260 - INFORMÁTICA Y SISTEMAS*
Ámbito: *075 - Ciencia De La Comp. E Intel. Artificial*
Área: *075 - Ciencia De La Comp. E Intel. Artificial*
Despacho: *INFORMÁTICA Y SISTEMAS*
Teléfono: *928458752* **Correo Electrónico:** *josejuan.hernandez@ulpgc.es*

Dr./Dra. José Évora Gómez
Departamento: *260 - INFORMÁTICA Y SISTEMAS*
Ámbito: *570 - Lenguajes Y Sistemas Informáticos*
Área: *570 - Lenguajes Y Sistemas Informáticos*
Despacho: *INFORMÁTICA Y SISTEMAS*
Teléfono: *928458728* **Correo Electrónico:** *jose.evora@ulpgc.es*

BIBLIOGRAFÍA

[1 Básico] Designing data-intensive applications :the big ideas behind reliable, scalable, and maintainable systems /

Martin Kleppmann.
O'Reilly,, Beijing, [China] : (2017)
9781449373320

[2 Recomendado] Big Data :conceptos, tecnologías y aplicaciones /

David Ríos Insúa, David Gómez-Ullate Oteiza.
Libros la Catarata :, Madrid : (2019)
978-84-00-10534-1

[3 Recomendado] Big data: análisis de grandes volúmenes de datos en organizaciones /

Luis Joyanes Aguilar.
Marcombo,, [Barcelona] : (2014)
9788426720818

[4 Recomendado] Practical big data analyticshands-on techniques to implement enterprise analytics and machine learning using Hadoop, Spark, NoSQL and R. /

Nataraj Dasgupta.
Packt Publishing,, Birmingham : (2018)
9781783554393

[5 Recomendado] Big data architect's handbook :a guide to building proficiency in tools and systems used by leading big data experts /

Syed Muhammad Fahad Akhtar.
Packt Publishing,, Birmingham, UK : (2018)
9781788835824