**Teaching Guide**

| Identifying Data | | | | 2022/23 |
|---|---|---|---|---|
| Subject (*) | Parallel Processing | | Code | 614G02023 |
| Study programme | Grao en Ciencia e Enxeñaría de Datos | | | |

| Descriptors | | | | |
|---|---|---|---|---|
| Cycle | Period | Year | Type | Credits |
| Graduate | 1st four-month period | Third | Obligatory | 6 |

| | |
|---|---|
| Language | Spanish |
| Teaching method | Face-to-face |
| Prerequisites | |
| Department | Enxeñaría de Computadores |
| Coordinador | Enes Álvarez, Jonatan | E-mail | jonatan.enes@udc.es |
| Lecturers | Enes Álvarez, Jonatan | E-mail | jonatan.enes@udc.es |
| Web | |
| General description | In this subject, the student will learn the basic role that the use of parallelism plays when it comes to accelerating the execution of programs in general, and of massive data processing in particular. The theory knowledge will start with the most basic concepts of parallelism, including its usefulness and applicability, the basic technical context of parallel programs (Chapter 1). Next, the main current hardware technologies that are used for parallel processing will be analyzed, includign their underlying technical details that allow to exploit parallelism out of programs (Chapter 2). After this, more advanced concepts regarding parallelism will be introduced, as well as classifications systems, software design patterns that allow to implement parallel programs, and techniques to measure the performance of such programs (Chapter 3). Finally, all of this acquired knowledge will be applied to study the state-of-the-art Big Data technologies and frameworks which allow for massive data processing (Chapter 4). On the more practical side of the subject, the student will complete several sessions with an incremental approach in order to gain the knowledge and ability to program and deploy solutions for the parallel processing of data. These sessions will start with a more technical, or 'low-level' approach, and will progress towards more complete, or 'high-level' solutions. These sessions will be also coordinated with the theory sessions so that any technique or technology has been previously introduced. In addition, these practical sessions will be self-contained and heavily focused to solving problems or scenarios with several specific techniques or technologies. This subject has a strong dependency with previous subjects like "Fundamentals of Programming I and Fundamentals of Programming II", mostly due to the technical programming ability, and with "Design and Analysis of Algorithms" for the knowledge to analyze the complexity of algorithms and programs. To a lesser extent, knowledge from the subject "Fundamentals of Computers" is advisable to understand the empirical behavior and the overall efficiency of some programs. |

| Study programme competences / results | |
|---|---|
| Code | Study programme competences / results |
| A12 | CE12 - Capacidade de coñecer e aplicar os principios fundamentais, principais paradigmas e técnicas da programación paralela e distribuída ao desenvolvemento de algoritmos para o procesamento e análise masiva de datos. |
| B2 | CB2 - Que os estudantes saiban aplicar os seus coñecementos ao seu traballo ou vocación dunha forma profesional e posúan as competencias que adoitan demostrarse por medio da elaboración e defensa de argumentos e a resolución de problemas dentro da súa área de estudo |
| B3 | CB3 - Que os estudantes teñan a capacidade de reunir e interpretar datos relevantes (normalmente dentro da súa área de estudo) para emitir xuízos que inclúan unha reflexión sobre temas relevantes de índole social, científica ou ética |
| B4 | CB4 - Que os estudantes poidan transmitir información, ideas, problemas e solucións a un público tanto especializado como non especializado |
| B7 | CG2 - Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables. |
| B8 | CG3 - Ser capaz de manter e estender formulacións teóricas fundadas para permitir a introdución e explotación de tecnoloxías novas e avanzadas no campo. |
| B9 | CG4 - Capacidade para abordar con éxito todas as etapas dun proxecto de datos: exploración previa dos datos, preprocesado, análise, visualización e comunicación de resultados. |

| B10 | CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións. |
|---|---|
| C1 | CT1 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida. |
| C4 | CT4 - Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade. |

| Learning outcomes | | | |
|---|---|---|---|
| **Learning outcomes** | **Study programme competences / results** | | |
| Know of and understand the technical requirements and the current technologies that allow for parallelism. | A12 | B8 B9 | |
| Know of the different currently available technologies to implement parallelism, their applicability, limits, advantages and disadvantages. | A12 | B2 B4 B8 B9 | |
| Be able to use parallelism techniques to adapt existing solutions so that they allow parallel processing. | A12 | B2 B4 B7 B8 B9 B10 | C1 |
| Be able to analyze the performance if a processing solution, with and without parallelization. | A12 | B2 B4 B7 B8 B9 B10 | C1 |
| Understand the paper that parallelization plays in today's society when it comes to key data processing tasks in business and research. | A12 | B3 B4 B8 B10 | C4 |

| Contents | |
|---|---|
| **Topic** | **Sub-topic** |
| Chapter 1 - Introduction and previous concepts | * The process and sequential program<br>* Lifecycle of a process<br>* Threads<br>* Paralell program<br>* Usefulness of parallelism |
| Chapter 2 - Hardware parellelism, hierarchy | * Levels of parallelism<br>* Internal processor parallelism (hidden)<br>* Processor functionalities (low-level parallelism)<br>* Processor accessible resources (high-level parallelism)<br>* Pool of machines (Cluster and Supercomputer)<br>* Distributed computing<br>* Specific devices<br>* State of the art of processors |

| Chapter 3 - Software parallelism, design and implementation | * Flynn taxonomy |
|---|---|
| | * Frameworks and languages for parallelism |
| | * Key concepts |
| | * Paradigms for parallel processing |
| | * Parallel programs analysis |
| | * Parallel programs design |
| Chapter 4 - Big Data technologies | * Data storage |
| | * Resource and execution management |
| | * Batch processing |
| | * Streaming processing |

| Planning | | | | |
|---|---|---|---|---|
| Methodologies / tests | Competencies / Results | Teaching hours (in-person & virtual) | Student?s personal work hours | Total hours |
| Guest lecture / keynote speech | A12 B3 B8 B9 C4 | 20 | 30 | 50 |
| Laboratory practice | A12 B2 B4 B7 B9 B10 C1 | 20 | 70 | 90 |
| Objective test | A12 B2 B4 B7 B9 C1 C4 | 3 | 1 | 4 |
| Personalized attention | | 6 | 0 | 6 |
| (*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students. | | | | |

| Methodologies | |
|---|---|
| Methodologies | Description |
| Guest lecture / keynote speech | * Theory sessions will introduce the basic knowledge later used on practice sessions.<br><br>* Other concepts will also be explained in detail, either because they are key to understand the technologies and techniques used on the practice sessions, or because they are more advanced and are crucial to understand the paper that parallelism has on nowadays society.<br>_____ |
| Laboratory practice | * Practice sessions will be self-contained and will deal with several specific problems or scenarios where parallelism plays an important role and where previously explained techniques or technologies are used.<br><br>* Each practice will focus on a single scenario or problem and will be composed of previous description and explanation, a proposed code to be analyzed and used, and a series of questions to work on. The student will have to work on the practice, starting on its first practice session and then continuing on its out-of-classroom time. The questions can range from performing an extension of the code, to performing an empirical study of its performance using several parallelism configurations, describing its behavior or functioning, or other types of questions overall focused at assessing the degree to which the student comprehended the problem and the solution.<br><br><br>* It is possible that for some practices, a brief quiz will be used. Nevertheless, such quiz will only be carried out once the practice has finished and submitted by all the students.<br>_____ |
| Objective test | * At the end of the term, and exam will be carried out to evaluate all the subject's knowledge, primarily the concepts from the theory sessions, but also to a lesser extent the ones from the practice sessions. |

| Personalized attention | |
|---|---|
| Methodologies | Description |

| Guest lecture / keynote speech Laboratory practice | * Personalized attention will focus on supporting the students with the overall subject.<br><br>* On the one hand, personalized attention will be available for those that have some issue understanding any concept exposed on the theory sessions, so that no student has any difficulty in keeping up with the classes and with those topics that will be the subject of evaluation.<br><br>* On the other hand, personalized attention will also be available for any student that requires some help with specific issues that arise from the practice lessons, whether they are due to technical problems or more deep understanding issues of the key concepts dealt with. Although this help will be available for any practice lesson throughout the term, it is advisable to deal with any doubt or problem either during the practice lesson or shortly afterwards.<br>_____<br><br>Those students with an approved dispensation for non-attendance at classes can also benefit by using this personalized attention to ask for the practice briefing as it was given during the ordinary practice classes. |
|---|---|

| Assessment | | | |
|---|---|---|---|
| **Methodologies** | **Competencies / Results** | **Description** | **Qualification** |
| Laboratory practice | A12 B2 B4 B7 B9 B10 C1 | * All the practice lessons will be the subject of evaluation and assessment by the teacher.<br><br>* Each practice lesson will be introduced and briefly explained by the teacher on its first associated practice class. The student is expected to start the practice lesson right away.<br><br>* The submission deadline of practice lessons will be previously agreed on, a time during which it is expected that the student carries out such practice lesson during the out-of-class time. The deadline will be group-specific.<br><br>* It is possible that for some practice lessons, the assessment score will be based partially or totally on a quiz that will be carried out on a date known beforehand. | 50 |
| Objective test | A12 B2 B4 B7 B9 C1 C4 | * Written exam carried out at the end of the term.<br><br>* It will mainly evaluate and assess concepts from the theory lessons.<br><br>* To a lesser point, some questions will also be present to re-asses key concepts from the practice lessons. | 50 |

| Assessment comments |
|---|
| In order to pass the subject, a minimum of 40% is required on the objective test, or final exam (2 points out of 5).Practice sessions will be NON REPEATABLE for the second chance.Part-time students can attend any practice class group, once it has been previously notified.Part-time students or students with approved dispensation for non-attendance at classes can submit their practice lessons taking into account the longest group-specific deadline available. In case a practice lesson is assessed using a quiz, a different date will be previously negotiated if needed. |

| Sources of information |
|---|

| Basic | - Francisco Almeida et al. (2008). Introducción a la programación paralela. Madrid : Paraninfo Cengage Learning |
|---|---|
| | - Jesús Carretero Pérez et al. (2007). Sistemas operativos : una visión aplicada . Madrid : McGraw-Hill |
| | - Giancarlo Zaccone (2015). Python Parallel Programming Cookbook. Packt Publishing |
| | - Tomasz Drabas, Denny Lee (2017). Learning PySpark: Build data-intensive applications locally and deploy at scale using the combined powers of Python and Spark 2.0. Packt Publishing |
| Complementary | - Bertil Schmidt et al. (2017). Parallel programming : concepts and practic. Cambridge, MA : Morgan Kaufmann |
| | - Peter S.Pacheco (2011). An introduction to parallel programming. Burlington, MA : Morgan Kaufmann |
| | -  Wes McKinney (2011). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly |

| Recommendations |
|---|
| **Subjects that it is recommended to have taken before** |
| Design and Analysis of Algorithms/614G02011 |
| Fundamentals of Computers/614G02005 |
| Fundamentals of Programming II/614G02009 |
| Fundamentals of Programming I/614G02004 |
| **Subjects that are recommended to be taken simultaneously** |
| |
| **Subjects that continue the syllabus** |
| Advanced Parallel Processing /614G02034 |
| **Other comments** |
| It is recommended to have some knowledge and ability to program with Python, as all it will be the language used for all of the practice lessons.It is recommended to have some degree of expertise with a Linux operating system, mainly process and filesystem management. |