

Course guide

270423 - CAP - High Performance Computing

Last modified: 10/07/2024

Unit in charge: Barcelona School of Informatics
Teaching unit: 701 - DAC - Department of Computer Architecture.

Degree: BACHELOR'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2021). (Compulsory subject).

Academic year: 2024 **ECTS Credits:** 6.0 **Languages:** Catalan

LECTURER

Coordinating lecturer: JOSEP LLUÍS BERRAL GARCÍA

Others: Primer quadrimestre:
JOSEP LLUÍS BERRAL GARCÍA - 11, 12
JORDI TORRES VIÑALS - 11, 12

PRIOR SKILLS

Having studied the subjects of Computer Fundamentals, as well as Parallelism and Distributed Systems.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

CE05. To be able to analyze and evaluate the structure and architecture of computers, as well as the basic components that make them up.

CE06. To be able to identify the features, functionalities and structure of Operating Systems and to design and implement applications based on their services.

CE07. To interpret the characteristics, functionalities and structure of Distributed Systems, Computer Networks and the Internet and design and implement applications based on them.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CE11. To identify and apply the fundamental principles and basic techniques of parallel, concurrent, distributed and real-time programming.

CE19. To use current computer systems, including high-performance systems, for the processing of large volumes of data from the knowledge of its structure, operation and particularities.

Generical:

CG1. To ideate, draft, organize, plan and develop projects in the field of artificial intelligence.

CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.

CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

Transversal:

CT2. Sustainability and Social Commitment. To know and understand the complexity of economic and social phenomena typical of the welfare society; Be able to relate well-being to globalization and sustainability; Achieve skills to use in a balanced and compatible way the technique, the technology, the economy and the sustainability.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Basic:

CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of study.

TEACHING METHODOLOGY

The course is based on theory and face-to-face laboratory sessions. The theoretical sessions combine lectures and seminars by experts in the field, following the program set out in this study plan and based on the use of own material. During the sessions, dialogue and discussion are promoted in order to anticipate and consolidate the learning outcomes of the subject.

The laboratory sessions deal with the aspects related to the different technologies presented, and follow the same topics as the syllabus studies. These are hands-on practical sessions, using different computational resources in the Department of Computer Architecture and the Barcelona Supercomputing Center.

LEARNING OBJECTIVES OF THE SUBJECT

1. Understand the use of high-performance computing and middlewares for artificial intelligence
2. Know the basic components of hardware and middleware in high-performance platforms
3. Learn about the use of accelerators (e.g. GPUs) and the tools for their exploitation
4. Learn about virtualization concepts and usage of virtual machines
5. Become familiar with the basic tools for exploiting distributed systems, with programming models oriented to distribution
6. Know the basic concepts on distributed systems, interconnection and connection among systems.
7. Learn about file systems: basic usage of file systems, redundancy on disks, logic volumes and fault tolerance.
8. Discover the challenges on high-performance computing on artificial intelligence

STUDY LOAD

Type	Hours	Percentage
Hours large group	30,0	20.00
Hours small group	30,0	20.00
Self study	90,0	60.00

Total learning time: 150 h

CONTENTS

Introduction to High-Performance Computing Systems

Description:

Introduction to large-scale computing systems, specialized and the Cloud.

Accelerators and high-performance devices

Description:

Incorporation of accelerators (e.g. GPUs) and the tools for their exploitation. Matrix operations accelerated through specialized devices.

Middleware and high-performance platforms for artificial intelligence

Description:

Basic components of hardware and middleware in high-performance platforms. Use of state of the art and commodity tools (e.g. TensorFlow, Pytorch, etc.) combined with specialized devices.



Parallelism applied to artificial intelligence

Description:

Parallelism on high-performance computing through the most common middlewares for artificial intelligence, deep learning and transformers, and their associated techniques

Introduction to distributed programming models for Big Data

Description:

Introduction to Map-Reduce programming models over distributed data systems and language Scala.

Virtualization concepts and containerization

Description:

Introduction to the use of virtual machines and containerization, for isolation executions and personalized environments, as load migration and resource management in shared systems.

Local and distributed file systems, redundancy and availability

Description:

Basic usage of file systems, distributed file systems, logic volumes, redundancy, fault tolerance and high availability.

Distributed systems for computing

Description:

Basic concepts on distributed systems (e.g. Hadoop and Spark), interconnection and communications, paradigms of distributed systems and protocols, and fault tolerance. Basic tools for exploiting concurrency on distributed systems, and their programming models oriented towards artificial intelligence and Big Data processing.

Challenges for high-performance computing for artificial intelligence

Description:

Challenges for present and future of high-performance computing applied to artificial intelligence. Current tools and environments in the industry, the Cloud, academia and society.

ACTIVITIES

Virtualization and containerization concepts

Description:

Introduction to the use of virtual machines and containerization, for isolated and customized execution of environments, as well as load migration and resource management to shared systems.

Specific objectives:

4

Related competencies :

CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

CE05. To be able to analyze and evaluate the structure and architecture of computers, as well as the basic components that make them up.

CE06. To be able to identify the features, functionalities and structure of Operating Systems and to design and implement applications based on their services.

CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.

CT2. Sustainability and Social Commitment. To know and understand the complexity of economic and social phenomena typical of the welfare society; Be able to relate well-being to globalization and sustainability; Achieve skills to use in a balanced and compatible way the technique, the technology, the economy and the sustainability.

Full-or-part-time: 20h

Theory classes: 4h

Laboratory classes: 4h

Self study: 12h

Service and application architecture

Description:

Introduction to Client-Server models, execution management systems, and launch of applications in cluster and Cloud systems.

Full-or-part-time: 10h

Theory classes: 2h

Laboratory classes: 2h

Self study: 6h

Supercomputers and High-performance Computing

Description:

Supercomputers and High Performance Computing systems, tools and environments. Familiarization with HPC facilities, hands-on use of HPC systems and C language.

Specific objectives:

2

Related competencies :

CE05. To be able to analyze and evaluate the structure and architecture of computers, as well as the basic components that make them up.

CE19. To use current computer systems, including high-performance systems, for the processing of large volumes of data from the knowledge of its structure, operation and particularities.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

CT2. Sustainability and Social Commitment. To know and understand the complexity of economic and social phenomena typical of the welfare society; Be able to relate well-being to globalization and sustainability; Achieve skills to use in a balanced and compatible way the technique, the technology, the economy and the sustainability.

Full-or-part-time: 13h

Theory classes: 2h

Laboratory classes: 2h

Self study: 9h

Accelerators, supercomputers and high-performance devices

Description:

Accelerators and high performance devices. GPUs and accelerator devices. Matrix multiplication using GPUs. Introduction to Python on a supercomputer.

Specific objectives:

3

Related competencies :

CE19. To use current computer systems, including high-performance systems, for the processing of large volumes of data from the knowledge of its structure, operation and particularities.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 13h

Theory classes: 2h

Laboratory classes: 2h

Self study: 9h

Computing in distributed systems

Description:

Basic concepts of distributed systems (e.g. Hadoop and Spark), interconnection and communications, distributed systems paradigms and protocols, and fault tolerance. Basic tools for the exploitation of concurrency in distributed systems, and their programming models oriented to artificial intelligence and massive data processing.

Specific objectives:

5, 6

Related competencies :

CE07. To interpret the characteristics, functionalities and structure of Distributed Systems, Computer Networks and the Internet and design and implement applications based on them.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CE11. To identify and apply the fundamental principles and basic techniques of parallel, concurrent, distributed and real-time programming.

CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

Full-or-part-time: 20h

Theory classes: 4h

Laboratory classes: 4h

Self study: 12h

Current tools and environments in industry, the cloud, academia and society.

Description:

Current tools and environments in industry, the cloud, academia and society.

Specific objectives:

8

Related competencies :

CG1. To ideate, draft, organize, plan and develop projects in the field of artificial intelligence.

CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence.

Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

CT2. Sustainability and Social Commitment. To know and understand the complexity of economic and social phenomena typical of the welfare society; Be able to relate well-being to globalization and sustainability; Achieve skills to use in a balanced and compatible way the technique, the technology, the economy and the sustainability.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

Full-or-part-time: 10h

Theory classes: 2h

Laboratory classes: 2h

Self study: 6h

Local and distributed file systems, redundancy and availability

Description:

Basic uses of file systems, as well as distributed data storage systems, logical volumes, redundancy, fault tolerance, and high availability.

Specific objectives:

7

Related competencies :

CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

CE07. To interpret the characteristics, functionalities and structure of Distributed Systems, Computer Networks and the Internet and design and implement applications based on them.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CE06. To be able to identify the features, functionalities and structure of Operating Systems and to design and implement applications based on their services.

CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 10h

Theory classes: 2h

Laboratory classes: 2h

Self study: 6h

Parallelism applied to artificial intelligence

Description:

Parallelism applied to artificial intelligence. Scalability, advanced deep learning techniques, transformers and the future of Deep Learning.

Specific objectives:

1, 2

Related competencies :

CE05. To be able to analyze and evaluate the structure and architecture of computers, as well as the basic components that make them up.

CE19. To use current computer systems, including high-performance systems, for the processing of large volumes of data from the knowledge of its structure, operation and particularities.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CG1. To ideate, draft, organize, plan and develop projects in the field of artificial intelligence.

CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence.

Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

CT2. Sustainability and Social Commitment. To know and understand the complexity of economic and social phenomena typical of the welfare society; Be able to relate well-being to globalization and sustainability; Achieve skills to use in a balanced and compatible way the technique, the technology, the economy and the sustainability.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

Full-or-part-time: 20h

Theory classes: 4h

Laboratory classes: 4h

Self study: 12h

Middleware and high-performance platforms for artificial intelligence

Description:

Middleware and high performance platforms for artificial intelligence. TensorFlow/Pytorch, Deep Learning, LLMs and HPC.

Specific objectives:

1

Related competencies :

CE19. To use current computer systems, including high-performance systems, for the processing of large volumes of data from the knowledge of its structure, operation and particularities.

CG1. To ideate, draft, organize, plan and develop projects in the field of artificial intelligence.

CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence.

Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

Full-or-part-time: 12h

Theory classes: 2h

Laboratory classes: 4h

Self study: 6h

Present and future challenges of high-performance computing applied to artificial intelligence. Seminars on HPC

Description:

Seminars of experts in the field. Presentation of work.

Specific objectives:

1, 5, 8

Related competencies :

CE07. To interpret the characteristics, functionalities and structure of Distributed Systems, Computer Networks and the Internet and design and implement applications based on them.

CE19. To use current computer systems, including high-performance systems, for the processing of large volumes of data from the knowledge of its structure, operation and particularities.

CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.

CE11. To identify and apply the fundamental principles and basic techniques of parallel, concurrent, distributed and real-time programming.

CG1. To ideate, draft, organize, plan and develop projects in the field of artificial intelligence.

CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.

CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

CT2. Sustainability and Social Commitment. To know and understand the complexity of economic and social phenomena typical of the welfare society; Be able to relate well-being to globalization and sustainability; Achieve skills to use in a balanced and compatible way the technique, the technology, the economy and the sustainability.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

Full-or-part-time: 22h

Theory classes: 6h

Laboratory classes: 4h

Self study: 12h

GRADING SYSTEM

The evaluation will basically be based on the completion of continuous work during the different sessions of the course. Attendance and participation will be mandatory, and therefore will also be assessed by passing a list and requiring participation in the interactive sessions. Finally, there will be a research project throughout the course, which students will have to present to their peers.

The distribution of weights for each activity is as follows:

- AS: attendance in class, theory and laboratories (10%), which will be used to evaluate transversal competence CT3.
- PR: class participation (10%)
- EX: laboratory deliverables (65%), as an arithmetic average of laboratory practices.
- RE: presentation of a research paper (15%), which will be used to evaluate transversal skills CT2, CT3 and CT6.

The Final Grade (NF) of the subject is obtained from

$$NF = 0.10 \times AS + 0.10 \times PR + 0.65 \times EX + 0.10 \times RE$$

BIBLIOGRAPHY

Basic:

- Torres, Jordi. First contact with Deep learning : practical introduction with Keras. Barcelona: Kindle Direct Publishing, [2018]. ISBN 9781983211553.
- Zhang, Aston and Lipton, Zachary C. and Li, Mu and Smola, Alexander J.. Dive into Deep Learning [on line]. [Berkeley]: The authors, 2020 [Consultation: 11/04/2025]. Available on: https://discovery.upc.edu/discovery/fulldisplay?docid=alma991001710129706711&context=U&vid=34CSUC_UPC:VU1&lang=ca.
- Sterling, Thomas; Anderson, Matthew; Brodowicz, Maciej. High performance computing : modern systems and practices. Cambridge, MA: Morgan Kaufmann, [2018]. ISBN 9780124201583.
- Chambers, B.; Zaharia, M. Spark: the definitive guide: big data processing made simple. Sebastopol, CA: O'Reilly, 2018. ISBN 9781491912300.
- White, Tom. Hadoop : the definitive guide. 4th ed. Sebastopol, Calif: O'Reilly, 2015. ISBN 9781491901632.
- Torres, J. La inteligencia artificial explicada a los humanos. Barcelona: Plataforma Editorial, 2023. ISBN 9788419655561.

Complementary:

- Barcelona Supercomputing Center. BSC documentation about Marenostrum 4 and CTE-Power. 2023.