Course guide
270649 - SA - Supercomputers Architecture

Last modified: 13/07/2022

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

Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).

Academic year: 2022 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer: JORDI TORRES VIÑALS

Others: Primer quadrimestre:
JORDI TORRES VIÑALS - 10

PRIOR SKILLS

Programming in C and Linux basics will be expected in the course. In addition, prior exposure to parallel programming constructions, Python language, experience with linear algebra/matrices, or machine learning knowledge will be helpful.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEE4.1. Capability to analyze, evaluate and design computers and to propose new techniques for improvement in its architecture.
CEE4.2. Capability to analyze, evaluate, design and optimize software considering the architecture and to propose new optimization techniques.
CEE4.3. Capability to analyze, evaluate, design and manage system software in supercomputing environments.

Generical:
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.

Transversal:
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.

Basic:
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

TEACHING METHODOLOGY

Class attendance and participation: Regular attendance is expected, and is required to be able to discuss concepts that will be covered during class.

Lab activities: Some exercises will be conducted as hands-on sessions during the course using supercomputing facilities. The student's own laptop will be required to access these resources during the theory class. Each hands-on session will involve writing a lab report with all the results. There are no days for theory classes and days for laboratory classes. Theoretical and practical activities will be interspersed during the same session to facilitate the learning process.

Reading/presentation assignments: Some exercise assignments will consist of reading documentation/papers that expand the concepts introduced during lectures. Some exercises will involve student presentations (randomly chosen).

Assessment: There will be one midterm exam in the middle of the course. The student is allowed to use any type of documentation (also digital via the student's laptop)

LEARNING OBJECTIVES OF THE SUBJECT

1.To train students to follow by themselves the continuous development of supercomputing systems that enable the convergence of advanced analytic algorithms as artificial intelligence.

STUDY LOAD

Type	Hours	Percentage
Hours large group	24,0	16.00
Hours small group	24,0	16.00
Self study	96,0	64.00
Guided activities	6,0	4.00

Total learning time: 150 h

CONTENTS

00. Welcome: Course content and motivation

01. Supercomputing basics

02. General purpose supercomputers

03. Parallel programming models

04. Parallel performance metrics

05. Parallel Performance models

06. Heterogeneous supercomputers

07. Parallel programming languages for heterogeneous platforms

08. Emerging Trends and Challenges in Supercomputing

09. Artificial Intelligence is a computing problem

10. Deep Learning essential concepts

11. Using Supercomputers for DL training

12. Accelerate the learning with parallel training using a multi-GPU parallel server

13. Accelerate the learning with parallel training using a multi-GPU parallel server

14. How to speed up the training of Transformers-based models

ACTIVITIES

00. Welcome

Specific objectives:
1

Related competencies :
CG1. Capability to apply the scientific method to study and analyse of phenomena and systems in any area of Computer Science, and in the conception, design and implementation of innovative and original solutions.
CEE4.3. Capability to analyze, evaluate, design and manage system software in supercomputing environments.
CEE4.1. Capability to analyze, evaluate and design computers and to propose new techniques for improvement in its architecture.
CEE4.2. Capability to analyze, evaluate, design and optimize software considering the architecture and to propose new optimization techniques.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.

Full-or-part-time: 2h 30m
Theory classes: 0h 30m
Self study: 2h

01. Supercomputing basics

Full-or-part-time: 5h 06m
Theory classes: 1h
Guided activities: 0h 06m
Self study: 4h

Exercise 01: Read and present a paper about exascale computers challenges

Full-or-part-time: 3h
Laboratory classes: 1h
Self study: 2h

02. General purpose supercomputers

Full-or-part-time: 5h
Theory classes: 1h
Self study: 4h

Exercise 02: Getting started with Supercomputing

Full-or-part-time: 3h 12m
Laboratory classes: 1h
Guided activities: 0h 12m
Self study: 2h

03. Parallel programming models

Full-or-part-time: 4h
Theory classes: 2h
Self study: 2h

Exercise 03: Getting Started with Parallel Programming Models

Full-or-part-time: 4h 06m
Laboratory classes: 2h
Guided activities: 0h 06m
Self study: 2h

04. Parallel performance metrics

Full-or-part-time: 6h
Theory classes: 2h
Self study: 4h

Exercise 04: Getting Started with Parallel Performance Metrics

Full-or-part-time: 5h 06m
Laboratory classes: 1h
Guided activities: 0h 06m
Self study: 4h

05. Parallel performance models

Full-or-part-time: 5h
Theory classes: 1h
Self study: 4h

Exercise 05: Getting started with parallel performance metrics and models

Full-or-part-time: 4h 06m
Laboratory classes: 1h
Guided activities: 0h 06m
Self study: 3h

06. Heterogeneous supercomputers

Full-or-part-time: 9h
Theory classes: 6h
Self study: 3h

Exercise 06: Comparing supercomputers performance

Full-or-part-time: 4h 06m
Laboratory classes: 1h
Guided activities: 0h 06m
Self study: 3h

07. Parallel programming languages for heterogeneous platforms

Full-or-part-time: 3h
Theory classes: 1h
Self study: 2h

Exercise 07: Getting started with CUDA

Full-or-part-time: 8h 06m
Laboratory classes: 3h
Guided activities: 0h 06m
Self study: 5h

08. Emerging Trends and Challenges in Supercomputing

Full-or-part-time: 2h
Theory classes: 1h
Self study: 1h

Exercise 08: Read and present a paper about emerging trends in supercomputing

Full-or-part-time: 4h 06m
Laboratory classes: 1h
Guided activities: 0h 06m
Self study: 3h

Midterm

Full-or-part-time: 12h
Theory classes: 2h
Self study: 10h

09. Artificial Intelligence is a Supercomputing problem

Full-or-part-time: 5h
Theory classes: 2h
Self study: 3h

Exercise 09: First contact with Deep Learning

Full-or-part-time: 6h 06m
Laboratory classes: 2h
Guided activities: 0h 06m
Self study: 4h

10. Deep Learning essential concepts

Full-or-part-time: 2h
Theory classes: 1h
Self study: 1h

Exercise 10: The new edition of the TOP500

Full-or-part-time: 5h 12m
Laboratory classes: 1h
Guided activities: 0h 12m
Self study: 4h

11. Using Supercomputers for DL training

Full-or-part-time: 4h
Theory classes: 2h
Self study: 2h

Exercise 11: Using a supercomputer for Deep Learning training

Full-or-part-time: 7h 12m
Laboratory classes: 3h
Guided activities: 0h 12m
Self study: 4h

12. Accelerate the learning with parallel training using a multi-GPU parallel server

Full-or-part-time: 2h
Theory classes: 1h
Self study: 1h

Exercise 12: Accelerate the learning with parallel training using a multi-GPU parallel server

Full-or-part-time: 7h 12m
Laboratory classes: 3h
Guided activities: 0h 12m
Self study: 4h

13. Accelerate the learning with distributed training using multiple parallel servers

Full-or-part-time: 2h
Theory classes: 1h
Self study: 1h

Exercise 13: Accelerate the learning with distributed training using multiple parallel server

Full-or-part-time: 11h 12m
Laboratory classes: 3h
Guided activities: 0h 12m
Self study: 8h

14. How to speed up the training of Transformers-based models

Full-or-part-time: 2h
Theory classes: 1h
Self study: 1h

Exercise 14: How to speed up the training of Transformers-based models

Full-or-part-time: 7h 12m
Laboratory classes: 3h
Guided activities: 0h 12m
Self study: 4h

Final remarks

Full-or-part-time: 2h 30m
Theory classes: 0h 30m
Self study: 2h

GRADING SYSTEM

The evaluation of this course can be obtained by continuous assessment. This assessment will take into account the following:

20% Attendance + participation
15% Midterm exam
65% Exercises (+ exercise presentations) and Lab exercises (+ Lab reports)
Details of the weight of each component of the course in the grade are described in the tentative scheduling section.

Course Exam: For those students who have not benefited from the continuous assessment, a course exam will be announced during the course. This exam includes evaluating the knowledge of the entire course (practical part, theoretical part, and self-learning part). During this exam, the student is not allowed to use any documentation (neither on paper nor digital).

BIBLIOGRAPHY

Basic:
- Torres, J. Class handouts and materials associated with this class. 2019.
- Torres, J. Understanding Supercomputing, to speed up machine learning algorithms (Course notes). 2018.
- BSC documentation. Marenostrum4 User's guide. Operations department, 2019.
- Sterling, T.; Anderson, M.; Brodowicz, M. High performance computing : modern systems and practices. Morgan Kaufmann, 2018. ISBN 9780124201583.
- Zhang, A.; Lipton, Z.C.; Li, M.; Smola, A.J. Dive into deep learning. 2020.
- Torres, J. First contact with Deep learning: practical introduction with Keras. Barcelona: Kindle Direct Publishing, 2018. ISBN 9781983211553.

RESOURCES

Hyperlink:
- https://torres.ai/SA-MIRI/