270647 - SCA - Supercomputing for Challenging Applications

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 701 - AC - Department of Computer Architecture
Academic year: 2018
Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6  Teaching languages: Catalan

Prior skills

Basic understanding of parallel architectures, including shared- and distributed-memory multiprocessor systems.

Useful programming skills of some parallel programming model.

Degree competences to which the subject contributes

Specific:

CEC3. Ability to apply innovative solutions and make progress in the knowledge that exploit the new paradigms of Informatics, particularly in distributed environments.
CEE3.3. Capability to understand the computational requirements of problems from non-informatics disciplines and to make significant contributions in multidisciplinary teams that use computing.
CEE4.2. Capability to analyze, evaluate, design and optimize software considering the architecture and to propose new optimization techniques.

General:

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.
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CG5. Capability to apply innovative solutions and make progress in the knowledge to exploit the new paradigms of computing, particularly in distributed environments.

Teaching methodology

During the course there will be two types of activities:

a) Activities focused on the acquisition of theoretical knowledge.
b) Activities focused on the acquisition of knowledge through experimentation by implementing and evaluating empirically in the laboratory the mechanisms explained at a theoretical level.

The theoretical activities include participatory lecture classes, which explain the basic contents of the course. The practical activities include seminar laboratories using the student's laptop in class, where students implement the mechanisms described in the lectures. The seminars require a preparation by reading the statement and supporting documentation, and a further elaboration of the conclusions in a report.

Learning objectives of the subject

1. The student should be able to understand the complexity of different algorithms, identify the computationally intensive parts of a simulation or data processing, and decide which parts need to be optimized and parallelized.
2. The student must be able to design and implement efficient parallel simulation and data processing algorithms using a parallel programming model.
3. The student must be able to evaluate the different tradeoffs (robustness, computational cost, scalability) in order to
select a specific algorithm for a simulation or data processing problem

<table>
<thead>
<tr>
<th>Study load</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total learning time:</strong> 150h</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
## Content

### Introduction

**Degree competences to which the content contributes:**

**Description:**
- Challenges in Science and Engineering
- HPC = Algorithms + Architecture + Programming Model
- Numerical and Non-Numerical Applications
- Parallel Computers
- Parallel Programming Models

### Introduction to Numerical Simulations

**Degree competences to which the content contributes:**

**Description:**
- From Models to Algorithms
- Discretization and PDEs
- Finite differences and Finite Elements
- Types of PDEs: Elliptic, Parabolic, Hyperbolic
- Initial-value and Boundary-value Problems
- Numerical schemes: Explicit vs. Implicit
- Stencils: Common Patterns
- Sparse Matrices and their Applications
- Numerical Software: From BLAS and LAPACK to Trilinos and PETSC

### Solving Large-Scale Linear Systems of Equations

**Degree competences to which the content contributes:**

**Description:**
- Direct vs Iterative Methods
- Fundamental Inner kernels and Matrix Formats
- Preconditioning
- Multigrid Methods
- Hierarchy of discretizations
- V-cycle iterations
- Partitioning and Reordering
- Local and global approaches: From RCM and MMD to ParMetis and PT-SCOTCH

### Practical Cases

**Degree competences to which the content contributes:**

**Description:**
- Computational Fluid Dynamics (CFD)
- Wave Propagation
# Introduction to the kernel of DBMSs and the execution of queries in such systems

**Degree competences to which the content contributes:**

**Description:**
This topic has the objective to understand the different software layers of a DBMS and how they interact, the different complexities that they incarnate and how their interaction determines the performance of such DBMSs. The sessions will be including discussions and the preparation of papers by the students.

---

# Big data kernels and graph databases

**Degree competences to which the content contributes:**

**Description:**
This session has the objective to introduce the students to the design of such Big-data and graph database systems, how they are built and how their performance is influenced by the structure of their different software layers. This sessions will be the base for those students who select this part of the course for the practical assignment.

---

# Benchmarking for Databases

**Degree competences to which the content contributes:**

**Description:**
Benchmarking is one of the most important issues in Database design and evolution. This part of the course will be designed to understand the different efforts being carried in benchmarking from the USA and Europe for relational and graph databases.

---

# Sequence Alignment

**Degree competences to which the content contributes:**

**Description:**
This is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Search, scoring and parallel strategies are important to overcome this challenge.

---

# Molecular dynamics

**Degree competences to which the content contributes:**

**Description:**
This consists in a computer simulation of physical movements of atoms and molecules. Cutoff techniques may be important to reduce the computing complexity of this challenge.
Protein-Protein Docking

Degree competences to which the content contributes:

Description:
This is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. There are serveral approaches that may increase the accuracy but also de complexity ot those methods.
### Planning of activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Hours</th>
<th>Description</th>
<th>Specific objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
<td>2h</td>
<td>Follow the lectures, study the materials and practices.</td>
<td>1</td>
</tr>
<tr>
<td><strong>Laboratory Environment and Brief explanation of the tools to be used</strong></td>
<td>14h</td>
<td>Follow the lectures, study the materials and practices.</td>
<td></td>
</tr>
<tr>
<td><strong>Part I: Numerical Applications</strong></td>
<td>48h</td>
<td>Follow the lectures, study the materials and practices.</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td><strong>Deliverable: Assignment on Numerical Applications</strong></td>
<td>0h</td>
<td>Assignment for the Numerical Applications part. To be delivered at Racó.</td>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>
## Part IIA) Non-Numerical Applications: Big-Data Management

**Hours:** 42h  
Theory classes: 18h  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 24h

**Description:**  
Follow the lectures, study the materials and practices.

**Specific objectives:**  
1, 2, 3

| Deliverable: Assignment on Big-Data Management | Hours: 0h  
Guided activities: 0h  
Self study: 0h |
|------------------------------------------------|----------------|
| **Description:**  
Assignment of the Big-Data Management module. To be delivered at Racó or by e-mail. | **Specific objectives:**  
1, 2, 3 |

## Part IIB) Non-Numerical Applications: Bio-Informatics

**Hours:** 44h  
Theory classes: 12h  
Practical classes: 0h  
Laboratory classes: 0h  
Guided activities: 0h  
Self study: 32h

**Description:**  
Follow the lectures, study the materials and practices.

**Specific objectives:**  
1, 2, 3

| Deliverable: Assignment on Bio-Informatics | Hours: 0h  
Guided activities: 0h  
Self study: 0h |
|--------------------------------------------|----------------|
| **Description:**  
Assignment of the Bio-Informatics module. To be delivered at Racó. | **Specific objectives:**  
1, 2, 3 |
Qualification system

The course will be evaluated with a set of assignments and a final project.

Grade = A_1/3 + A_2/3 + 0.A_3/3

Where

A_i := Assignment i (i from 1 to 3)
Bibliography

Basic:


Complementary:


Others resources:

Hyperlink

http://glaros.dtc.umn.edu/gkhome/views/metis/

http://www.open-mpi.org/

http://openmp.org/

http://www.netlib.org/utk/people/JackDongarra/la-sw.html

http://www.mgnet.org/

http://www.labri.fr/perso/pelegrin/scotch/