

Guía docente

270647 - SCA - Supercomputación para Aplicaciones Exigentes

Última modificación: 16/07/2024

Unidad responsable: Facultad de Informática de Barcelona

Unidad que imparte: 701 - DAC - Departamento de Arquitectura de Computadores.

Titulación: MÁSTER UNIVERSITARIO EN INNOVACIÓN E INVESTIGACIÓN EN INFORMÁTICA (Plan 2012). (Asignatura optativa).

Curso: 2024

Créditos ECTS: 6.0

Idiomas: Inglés

PROFESORADO

Profesorado responsable: CARLOS ALVAREZ MARTINEZ

Otros: Primer quadrimestre:
CARLOS ALVAREZ MARTINEZ - 10
JOSE RAMON HERRERO ZARAGOZA - 10
JOSEP LARRIBA PEY - 10

CAPACIDADES PREVIAS

Comprensión básica de las arquitecturas paralelas, incluidos los sistemas multiprocesador de memoria compartida y distribuida.

Habilidades de programación útiles de algún modelo de programación paralela.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CEE4.2. Capacidad de analizar, evaluar, diseñar y optimizar software considerando la arquitectura y de proponer nuevas técnicas de optimización.

Genéricas:

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

CG5. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

METODOLOGÍAS DOCENTES

Durante el curso se realizarán dos tipos de actividades:

- Actividades centradas en la adquisición de conocimientos teóricos.
- Actividades centradas en la adquisición de conocimientos mediante experimentación con la implementación y evaluación empírica en el laboratorio de los mecanismos explicados a nivel teórico.

Las actividades teóricas incluyen clases expositivas participativas donde se explican los contenidos básicos del curso. Las actividades prácticas incluyen seminarios de laboratorio usando el ordenador portátil del alumno/a, donde los alumnos implementan los mecanismos descritos en las clases expositivas. Los seminarios requieren de una preparación previa mediante la lectura del enunciado y la documentación de apoyo, y una elaboración posterior de las conclusiones obtenidas en un informe.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

- El estudiante debe ser capaz de comprender la complejidad de los diferentes algoritmos, identificar las partes computacionalmente intensivas de una simulación o procesamiento de datos, y decidir qué partes necesitan optimizarse y paralelizarse.
- El estudiante debe ser capaz de diseñar e implementar algoritmos eficientes de simulación paralela y procesamiento de datos usando un modelo de programación paralelo.
- El estudiante debe ser capaz de evaluar los diferentes compromisos (robustez, coste computacional, escalabilidad) para seleccionar un algoritmo específico para un problema de simulación o procesamiento de datos

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTE

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	96,0	64.00
Horas grupo grande	54,0	36.00

Dedicación total: 150 h

CONTENIDOS

Introducción

Descripción:

- Introducción: descripción general
- El método científico moderno
- Simulación y optimización
- HPC vs. HTC
- Simulaciones numéricas
- Límites de la paralelización
- Evolución y límites de los sistemas HPC

Introducción a las Simulaciones Numéricas

Descripción:

- De modelos a algoritmos
- Discretización y PDE
- Tipo de PDEs: elípticas, parabólicas, hiperbólicas
- Del problema a las matemáticas y la solución
- Esquemas numéricos: Explicit vs Implicit
- Diferencias finitas y elementos finitos

Resolución directa de sistemas lineales de ecuaciones a gran escala

Descripción:

- Sistemas triangulares y paralelización
- Eliminación de Gauss
- Factorización LU
- Métodos de particionamiento
- HPL

Resolución iterativa de sistemas lineales de ecuaciones a gran escala

Descripción:

- Métodos directos vs iterativos
- Jacobi
- Paralelización de métodos iterativos
- Gauss-Seidel y SOR
- Métodos de Krylov y preconditionamiento: HPCG
- Software para métodos numéricos: BLAS, LAPACK, etc.

Casos prácticos de sistemas numéricos

Descripción:

- Modelos de programación en paralelo
- Paralelismo y granularidad
- Paralelismo de bloques
- Resolución sistemas dispersos

Introducción a los problemas de ML

Descripción:

- Introducción a las operaciones comunes de las DNNs
- Aplicaciones de DNNs más comunes
- Arquitecturas para DNNs
- Optimizaciones para DNNs
- DNNs y precisión
- Paralelismo en DNNs

Alineación de secuencia

Descripción:

Se trata de una forma de ordenar las secuencias de ADN, ARN o proteína para identificar regiones de semejanza que pueden ser consecuencia de relaciones funcionales, estructurales o evolutivos entre las secuencias.

Las estrategias de búsqueda, puntuación y de paralelización son importantes para superar este reto.

Casos prácticos

Descripción:

- DNNs y precisión
- Paralelismo en DNNs
- Algoritmos de Sequence Alignment

Introducción al kernel de DBMSs y la ejecución de consultas en tales sistemas

Descripción:

Este tema tiene como objetivo comprender las diferentes capas de software de un DBMS y cómo interactúan, las diferentes complejidades que encarnan y cómo su interacción determina el rendimiento de estos DBMS.

Aplicaciones de Big Data y bases de datos de grafos

Descripción:

Esta sesión tiene el objetivo de presentarles a los estudiantes el diseño de tales sistemas de base de datos Big Data y de grafos, cómo se construyen y cómo su estructura se ve influenciada por la estructura de sus diferentes capas de software.

Benchmarking para bases de datos

Descripción:

El benchmarking es uno de los problemas más importantes en el diseño y la evolución de la base de datos. Esta parte del curso será diseñada para comprender los diferentes esfuerzos realizados en benchmarking de EEUU y Europa para bases de datos relacionales y grafos.

Casos prácticos de Big Data

Descripción:

Casos prácticos de Big Data

ACTIVIDADES

Introducción

Descripción:

Siga las conferencias, estudiar los materiales y las prácticas.

Objetivos específicos:

1

Competencias relacionadas:

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Dedicación: 2h

Grupo grande/Teoría: 2h

Parte I: Aplicaciones Numéricas

Descripción:

Siga las conferencias, estudie los materiales y las prácticas.

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CEE4.2. Capacidad de analizar, evaluar, diseñar y optimizar software considerando la arquitectura y de proponer nuevas técnicas de optimización.

CG5. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Dedicación: 18h

Aprendizaje autónomo: 10h

Grupo grande/Teoría: 8h

Laboratorio de aplicaciones numéricas

Dedicación: 22h

Aprendizaje autónomo: 12h

Actividades dirigidas: 2h

Grupo pequeño/Laboratorio: 8h

Entregable: asignación en aplicaciones numéricas

Descripción:

Asignación para la parte de Aplicaciones Numéricas. Para ser entregado en Racó.

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CEE4.2. Capacidad de analizar, evaluar, diseñar y optimizar software considerando la arquitectura y de proponer nuevas técnicas de optimización.

CG5. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Dedicación: 12h

Aprendizaje autónomo: 12h

Parte II: Aplicaciones no numéricas: ML y Bioinformática

Descripción:

Sigue las clases magistrales, estudia los materiales y prácticas

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CEE4.2. Capacidad de analizar, evaluar, diseñar y optimizar software considerando la arquitectura y de proponer nuevas técnicas de optimización.

CG5. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Dedicación: 20h

Aprendizaje autónomo: 10h

Grupo grande/Teoría: 10h

Laboratorio de aplicaciones no numéricas

Dedicación: 16h

Aprendizaje autónomo: 10h

Grupo pequeño/Laboratorio: 6h

Entregable: Problema no numérico

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CEE4.2. Capacidad de analizar, evaluar, diseñar y optimizar software considerando la arquitectura y de proponer nuevas técnicas de optimización.

CG5. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Dedicación: 12h

Aprendizaje autónomo: 12h

Part III: Problemas de Big Data

Descripción:

Siga las conferencias, estudie los materiales y las prácticas.

Dedicación: 20h

Aprendizaje autónomo: 10h

Grupo grande/Teoría: 10h

Laboratorio Big Data

Dedicación: 16h

Aprendizaje autónomo: 10h

Grupo pequeño/Laboratorio: 6h

Entregable de Big Data

Objetivos específicos:

1, 2, 3

Competencias relacionadas:

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEE3.3. Capacidad para entender las necesidades computacionales de problemas de disciplinas distintas de la informática y efectuar contribuciones significativas en equipos multidisciplinares que usen la computación.

CEE4.2. Capacidad de analizar, evaluar, diseñar y optimizar software considerando la arquitectura y de proponer nuevas técnicas de optimización.

CG5. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CG1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Dedicación: 12h

Aprendizaje autónomo: 12h

SISTEMA DE CALIFICACIÓN

El curso se evaluará con una nota parcial por bloque de contenidos (Numéricos, No numéricos y Big Data). Cada bloque pesará igual en la nota final:

$$\text{Nota} = A1 / 3 + A2 / 3 + A3 / 3$$

Dónde

A_i : = Nota del bloque i (i de 1 a 3)

y

- Cada parte del curso (bloque) tiene algunos entregables cortos y parciales (PD) y un entregable del proyecto final (FPD).

BIBLIOGRAFÍA

Básica:

- Burnett, D.S. Finite element analysis: from concepts to applications. Addison-Wesley, 1987. ISBN 0201108062.
- Chapra, S.C.; Canale, R.P. Numerical methods for engineers. 8th ed. New York: McGraw-Hill, 2021. ISBN 9781260571387.
- Pržulj, N. Analyzing network data in biology and medicine: an interdisciplinary textbook for biological, medical and computational scientists. Cambridge: Cambridge University Press, 2019. ISBN 9781108432238.
- Saad, Y. Iterative methods for sparse linear systems. 2nd ed. SIAM, 2003. ISBN 0898715342.
- Quinn, M.J. Parallel programming in C with MPI and OpenMP. International edition. McGraw-Hill, 2003. ISBN 0071232656.
- Briggs, W.L.; Henson, V.E.; McCormick, S.F. A multigrid tutorial. 2nd ed. Society for Industrial and Applied Mathematics, 2000. ISBN 0898714621.
- Tramontano, A. The ten most wanted solutions in protein bioinformatics. Chapman and Hall/CRC, 2005. ISBN 1584884916.
- Gray, J.; Reuter, A. Transaction processing: concepts and techniques. Morgan Kaufmann, 1993. ISBN 1558601902.

Complementaria:

- Cohen, G.C. Higher-order numerical methods for transient wave equations. Pržulj, Natasa. Springer, 2001. ISBN 354041598X.
- Fichtner, A. Full seismic waveform modelling and inversion. Springer, 2011. ISBN 9783642158063.
- Wesseling, P. An introduction to multigrid methods. Edwards, 2004. ISBN 1930217080.

RECURSOS

Enlace web:

- <http://glaros.dtc.umn.edu/gkhome/views/metis/>- <http://openmp.org/>- <http://www.mgnet.org/>- <http://www.netlib.org/utk/people/JackDongarra/la-sw.html>- <http://www.open-mpi.org/>