



# Guia docent

## 270218 - PSD - Paral·lelisme i Sistemes Distribuïts

Última modificació: 30/01/2024

**Unitat responsable:** Facultat d'Informàtica de Barcelona  
**Unitat que imparteix:** 701 - DAC - Departament d'Arquitectura de Computadors.  
**Titulació:** GRAU EN CIÈNCIA I ENGINYERIA DE DADES (Pla 2017). (Assignatura obligatòria).  
**Curs:** 2023      **Crèdits ECTS:** 6.0      **Idiomes:** Català

### PROFESSORAT

---

**Professorat responsable:** JULITA CORBALAN GONZALEZ  
**Altres:** Segon quadrimestre:  
YOLANDA BECERRA FONTAL - 11, 12, 13  
JULITA CORBALAN GONZALEZ - 11, 12, 13

### CAPACITATS PRÈVIES

---

C i Python són els llenguatges de programació triats per a les sessions de laboratori d'aquest curs. Se suposa que l'alumne té un coneixement bàsic de Python i C abans de començar les classes.

### COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

---

#### Específiques:

CE4. Utilitzar els sistemes de computació actuals, inclosos els sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.

#### Genèriques:

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CG4. Identificar oportunitats per a aplicacions innovadores orientades a dades en entorns tecnològics en contínua evolució.

#### Transversals:

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT5. Ús solvent dels recursos d'informació. Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació en l'àmbit de l'especialitat i valorar de forma crítica els resultats d'aquesta gestió.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

#### Bàsiques:

CB1. Que els estudiants hagin demostrat posseir i comprendre coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es recolza en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda del seu camp d'estudi.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

## METODOLOGIES DOCENTS

Durant el curs hi haurà quatre tipus d'activitats:

- Activitats orientades a l'adquisició de coneixements teòrics. Les activitats teòriques inclouen classes de classes participatives, que expliquen els continguts bàsics del curs.
- Les activitats es centren en l'adquisició del coneixement a través de l'experimentació mitjançant l'enfocament "aprenent per fer" en sessions de laboratori guiades per pràctiques (i informe final). Algunes sessions poden incloure treball previ o a fer després de les sessions depenent de l'aprofitament dels laboratoris.
- Poques sessions durant les classes teòriques on es realitzaran exercicis pràctics per realitzar avaluacions numèriques i anàlisi per a l'avaluació del rendiment
- Dos informes dels exercicis que es duran a terme als laboratoris relacionats amb entorns i aplicacions HPC i amb entorns d'anàlisi de dades

Aquest quadrimestre, ja que les classes de laboratori es faran en aules de teoria, els estudiants hauran de portar el seu propi portàtil. Per fer els examens, tant de teoria com de laboratori, degut a que s'entregaran en format digital, també serà necessari que portin el seu propi portàtil. Totes les classes de teoria que es facin online s'habilitarà un meet a l'horari oficial. Per les classes de laboratori, aquells alumnes que estiguin confinats s'habilitarà un meet per poder seguir les classes.

## OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

- Conèixer els fonaments dels sistemes paral·lels i distribuïts actuals
- Coneixer i saber usar els elements bàsics que conformen els sistemes paral·lels i distribuïts
- Coneixer i poder triar convenientment quin els entorns d'anàlisi avançada que usen sistemes distribuïts i paral·lel
- Us pràctic per diferents problemes plantejats dels entorns cloud, sistemes paral·lels i distribuïts disponibles actualment per a un enginyer i científic de dades
- Familiaritzar-se amb els models de programació més habituals dels sistemes paral·lels i distribuïts

## HORES TOTALS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup petit	30,0	20.00
Hores grup gran	30,0	20.00
Hores aprenentatge autònom	90,0	60.00

**Dedicació total:** 150 h

## CONTINGUTS

### Fonaments de supercomputació paral·lela i distribuïda

**Descripció:**

En aquest tema, els alumnes aprendran conceptes bàsics de la computació paral·lela així com mètriques que els ajudaran a avaluar tant el rendiment dels seus programes com els límits derivats de la pròpia estructura de l'aplicació.

### Arquitecturas paraleles i distribuïdes

**Descripció:**

En aquest tema, els estudiants aprenen les principals característiques de les arquitectures paral·leles i distribuïdes que poden influir en el disseny dels seus programes d'anàlisi de dades o d'entendre el rendiment (o pèrdua de rendiment) d'aquests.



### Entorns d'execució de computació paral·lela i anàlisi de dades

**Descripció:**

En aquest tema els alumnes aprendran els diferents entorns que principalment poden trobar-se al executar tants les aplicacions de generar dades com les que els emmagatzemen o analitzen. Es farà èmfasi en les diferències entre els tres entorns i en l'impacte que pot tenir en l'eficiència de les seves aplicacions.

### Models de programació per a supercomputadors

**Descripció:**

En aquest tema els alumnes veuran els principis bàsics dels models de programació més utilitzats en els entorns HPC: MPI, OpenMP i models híbrids MPI + OpenMP. Es donaran les eines per detectar i gestionar els principals detalls que puguin afectar tant la robustesa dels seus programes com a l'eficiència.

També s'introduiran models orientats a coprocessadors que ofereixin un bon rendiment vs. eficiència. Consum d'energia i molt usat en l'anàlisi de dades.

### New programari requirements for advanced analytics

**Descripció:**

En aquest tema, els estudiants veuran amb més detall les característiques dels models de programació i entorns d'execució per a l'emmagatzematge i l'anàlisi de dades. El model Apache Spark / Hadoop s'utilitzarà com a referència, com a referència per a l'emmagatzematge de dades Cassandra i com a eines d'anàlisi TensorFlow / keras.

### Machine Learning en Supercomputadors: Cas d'us basat en Spark/Cassandra/TensorFlow

**Descripció:**

En aquest tema aprofundirem en un entorn de Machine learning utilitzant com a referència el model Apache Spark, com DB key / value Cassandra i com a eina d'anàlisi TensorFlow. S'explicaran els elements més importants d'aquests tres components que puguin afectar en major mesura tant al disseny d'aplicacions de machine learning com l'emmagatzematge de dades i anàlisi.

### Sessions de laboratori

**Descripció:**

Les sessions de laboratori s'agruparan en dos projectes que s'aniran realitzant tant en les sessions de laboratori com en treball autònom. Els dos projectes estaran relacionats amb la programació, anàlisi i optimització d'un cas el més realista possible en dos entorns: entorns d'execució paral·lel (mpi + OpenMP, sistemes de cues, etc), utilitzats per generar i post-processar dades, i entorns específics de gestió i anàlisi de dades com és Apache Spark + Cassandra + TensorFlow.

## ACTIVITATS

### Introducció de l'assignatura

**Descripció:**

Durant aquesta activitat s'explicaran els objectius, continguts, i funcionament de l'assignatura

**Dedicació:** 1h

Grup gran/Teoria: 1h



### Desenvolupament del tema "Fonaments de supercomputació paral·lela i distribuïda"

**Descripció:**

En aquest tema, els alumnes aprendran conceptes bàsics de la computació paral·lela així com mètriques que els ajudaran a avaluar tant el rendiment dels seus programes com els límits derivats de la pròpia estructura de l'aplicació.

**Objectius específics:**

1

**Competències relacionades:**

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CB1. Que els estudiants hagin demostrat posseir i comprendre coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es recolza en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda del seu camp d'estudi.

**Dedicació:** 4h

Grup gran/Teoria: 2h

Aprenentatge autònom: 2h

### Desenvolupament del tema "Arquitecturas paraleles i distribuïdes"

**Descripció:**

En aquest tema, els estudiants aprenen les principals característiques de les arquitectures paral·leles i distribuïdes que poden influir-les en el disseny dels seus programes d'anàlisi de dades i entendre el rendiment (o pèrdua de rendiment) d'aquests : Es veuran, per exemple característiques de sistemes amb arquitectura multi-core, hyperthreading, memòria compartida-distribuïda, localitat de dades temporal-espacial, tipus storage (local, remot), tipologia xarxes, etc.

**Objectius específics:**

1, 2

**Competències relacionades:**

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

CB1. Que els estudiants hagin demostrat posseir i comprendre coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es recolza en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda del seu camp d'estudi.

**Dedicació:** 4h

Grup gran/Teoria: 2h

Aprenentatge autònom: 2h

### Desenvolupament del tema "Entorns d'execució de computació paral·lela i anàlisi de dades"

#### Descripció:

En aquest tema els alumnes aprendran els diferents entorns que principalment poden trobar-se al executar tants les aplicacions de generar dades com les que els emmagatzemen o analitzen. Es farà èmfasi en les diferències entre els tres entorns i en l'impacte que pot tenir en l'eficiència de les seves aplicacions. Entorno de ejecución con colas para HPC, cloud computing para DA. Durant aquest tema es dividirà en entorns HPC i entorns d'anàlisi de dades (DA). També es faran exercicis de problemes durant les classes de teoria.

#### Objectius específics:

2, 3

#### Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.  
CG4. Identificar oportunitats per a aplicacions innovadores orientades a dades en entorns tecnològics en contínua evolució.  
CE4. Utilitzar els sistemes de computació actuals, inclosos els sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.  
CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.  
CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.  
CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.  
CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

**Dedicació:** 12h

Grup gran/Teoria: 6h

Aprenentatge autònom: 6h

### Desenvolupament del tema "Models de programació per a supercomputadors"

#### Descripció:

En aquest tema els alumnes veuran els principi bàsics dels models de programació més utilitzats en els entorns HPC: MPI, OpenMP i models híbrids MPI + OpenMP. Es donaran les eines per detectar i gestionar els principals detalls que puguin afectar tant la robustesa dels seus programes com a l'eficiència. També s'introduiran models orientats a coprocessadors que ofereixin un bon rendiment vs. eficiència. Consum d'energia i molt usat en l'anàlisi de dades.

#### Objectius específics:

5

#### Competències relacionades:

CE4. Utilitzar els sistemes de computació actuals, inclosos els sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.  
CT5. Ús solvent dels recursos d'informació. Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació en l'àmbit de l'especialitat i valorar de forma crítica els resultats d'aquesta gestió.  
CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

**Dedicació:** 12h

Grup gran/Teoria: 6h

Aprenentatge autònom: 6h



### Desenvolupament del tema "New programari per anàlisi de dades"

**Descripció:**

En aquest tema, els estudiants veuran amb més detall les característiques dels models de programació i entorns d'execució per a l'emmagatzematge i l'anàlisi de dades. El model Apache Spark / Hadoop s'utilitzarà com a referència, com a referència per a l'emmagatzematge de dades Cassandra i com a eines d'anàlisi TensorFlow / keras.

**Objectius específics:**

3

**Competències relacionades:**

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.  
CG4. Identificar oportunitats per a aplicacions innovadores orientades a dades en entorns tecnològics en contínua evolució.  
CE4. Utilitzar els sistemes de computació actuals, inclosos els sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.

**Dedicació:** 14h

Grup gran/Teoria: 7h

Aprenentatge autònom: 7h

### Desenvolupament del tema "Machine Learning en Supercomputadors: Cas d'us basat en Spark/Cassandra/TensorFlow"

**Descripció:**

En aquest tema aprofundirem en un entorn de Machine learning utilitzant com a referència el model Apache Spark, com DB key / value Cassandra i com a eina d'anàlisi TensorFlow. S'explicaran els elements més importants d'aquests tres components que puguin afectar en major mesura tant al disseny d'aplicacions de machine learning com l'emmagatzematge de dades i anàlisi.

**Objectius específics:**

4

**Competències relacionades:**

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.  
CE4. Utilitzar els sistemes de computació actuals, inclosos els sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.  
CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.  
CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.  
CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

**Dedicació:** 8h

Grup gran/Teoria: 4h

Aprenentatge autònom: 4h



### Sesions de laboratori i entregables: Execució d'aplicacions HPC, Generació de dades en entorns HPC, enmagatzematge i anàlisi de dades en contexte de DA (Data Analytics)

#### Descripció:

Aquest projecte seguirà el fluxe natural de les dades, des de la paral·lelització, execució, i avaluació de codis que generen dades, tradicionalment en un entorn HPC, passant per l'enmagatzematge i el posterior anàlisi amb els nous entorns d'execució per aquest tipus de problemes.

#### Objectius específics:

2, 3, 4, 5

#### Competències relacionades:

CG2. Elegir i aplicar els mètodes i tècniques més adequats a un problema definit per dades que representin un repte pel seu volum, velocitat, varietat o heterogeneïtat, inclosos mètodes informàtics, matemàtics, estadístics i de processament del senyal.

CG4. Identificar oportunitats per a aplicacions innovadores orientades a dades en entorns tecnològics en contínua evolució.

CG1. Concebre sistemes computacionals que integren dades de procedències i formes molt diverses, construeixen amb ells models matemàtics, raonen sobre aquests models i actuen en conseqüència, aprenent de l'experiència.

CE4. Utilitzar els sistemes de computació actuals, inclosos els sistemes d'alt rendiment, per al procés de grans volums de dades des del coneixement de la seva estructura, funcionament i particularitats.

CT6. Aprenentatge autònom. Detectar deficiències en el propi coneixement i superar-les mitjançant la reflexió crítica i l'elecció de la millor actuació per ampliar aquest coneixement.

CT4. Treball en equip. Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o realitzant tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, assumint compromisos tenint en compte els recursos disponibles.

CT5. Ús solvent dels recursos d'informació. Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació en l'àmbit de l'especialitat i valorar de forma crítica els resultats d'aquesta gestió.

CT7. Tercera llengua. Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i d'acord amb les necessitats que tindran els titulats i titulades.

CB5. Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

CB2. Que els estudiants sàpiguen aplicar els seus coneixements al seu treball o vocació d'una manera professional i posseeixin les competències que solen demostrar-se mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins la seva àrea d'estudi.

#### Dedicació: 56h

Grup petit/Laboratori: 28h

Aprenentatge autònom: 28h

## SISTEMA DE QUALIFICACIÓ

- Exàmen parcial: 35% (Primera part del curs)
- Exàmen final: 35% (Segona part del curs)
- Laboratori: 30%.

La nota de laboratori sortirà de l'avaluació dels entregables de laboratori

La nota serà  $0.3 \cdot \text{lab} + 0.35 \cdot \text{final} + 0.35 \cdot \text{parcial}$ .

Reavaluació: Només aquells/es que hagin suspès podran presentar-se a l'examen de Reavaluació (examen només de teoria). En aquest cas, la nota final de l'assignatura serà:

Nota Final:  $\text{Maxim}(\text{Exàmen reavaluació} \cdot 0.7, \text{Examenparcial} \cdot 0.35 + \text{ExamenFinal} \cdot 0.35) + \text{laboratori} \cdot 0.3$



## BIBLIOGRAFIA

---

### **Bàsica:**

- TORRES, Jordi. Hand-on sessions at GitHub.
- Torres, J. Slides of the course. UPC,
- Torres, J. Understanding supercomputing: with Marenostum Supercomputer in Barcelona. Universitat Politècnica de Catalunya, Barcelona Supercomputing Center, 2016. ISBN 9781365376825.
- Torres, J. Hello world en TensorFlow. Universitat Politècnica de Catalunya, Barcelona Supercomputing Centre, 2016. ISBN 9781326532383.
- Macias, M.; Gómez, M.; Tous, R.; Torres, J. Introducción a Apache Spark: para empezar a programar el big data. UOC, 2015. ISBN 9788491160373.
- Articles from Technical Journals in the area.

### **Complementària:**

- Torres, J. Empresas en la nube: ventajas y retos del cloud computing. Libros de Cabecera, 2011. ISBN 9788493908225.