



| Guía Docente          |  |                    |  |          |
|-----------------------|--|--------------------|--|----------|
| Datos Identificativos |  |                    |  | 2023/24  |
| Asignatura (*)        | Análise de Datos con HPC   | Código             | 614473108  |          |
| Titulación            | Mestrado Universitario en Computación de Altas Prestacións / High Performance Computing (Mod. Presencial)  |                    |  |          |
| Descriptorios         |  |                    |  |          |
| Ciclo                 | Período  | Curso              | Tipo   | Créditos |
| Mestrado Oficial      | 2º cuatrimestre  | Primeiro           | Optativa   | 6        |
| Idioma                | Inglés   |                    |  |          |
| Modalidade docente    | Presencial   |                    |  |          |
| Prerrequisitos        |  |                    |  |          |
| Departamento          | Enxeñaría de Computadores  |                    |  |          |
| Coordinación          | López Taboada, Guillermo   | Correo electrónico | guillermo.lopez.taboada@udc.es                             |          |
| Profesorado           | López Taboada, Guillermo<br>Rodríguez Álvarez, Gabriel   | Correo electrónico | guillermo.lopez.taboada@udc.es<br>gabriel.rodriguez@udc.es |          |
| Web                   | aula.cesga.es  |                    |  |          |
| Descrición xeral      | A cantidade cada vez maior de información accesible a través de Internet fai que o procesamento eficiente de grandes cantidades de datos sexa cada vez de maior interese. Isto levou ao desenvolvemento de novas técnicas de almacenamento e procesamento de inxentes cantidades de información, denominadas técnicas Big Data, que se adaptan de forma natural aos sistemas distribuídos. |                    |  |          |

| Competencias / Resultados do título |   |
|-------------------------------------|---|
| Código                              | Competencias / Resultados do título   |
| A1                                  | CE1 - Definir, avaliar e seleccionar a arquitectura e o software máis axeitado para a resolución dun problema   |
| A2                                  | CE2 - Analizar e mellorar o rendimento dunha arquitectura ou un software dado   |
| B1                                  | CB6 - Posuir e comprender coñecementos que aporten unha base ou oportunidade de ser orixinais no desenrolo e/ou aplicación de ideas, a miúdo nun contexto de investigación  |
| B2                                  | CB7 - Que os estudantes saiban aplicar os coñecementos adquiridos e súa capacidade de resolución de problemas en contornas novas ou pouco coñecidas dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo |
| B6                                  | CG1 - Ser capaz de buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo   |
| B8                                  | CG3 - Ser capaz de manter e extender plantexamentos teóricos fundados para permitir a introducción e explotación de tecnoloxías novas e avanzadas no campo  |
| B10                                 | CG5 - Ser capaz de traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións.   |
| C1                                  | CT1 - Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.   |
| C4                                  | CT4 - Valorar a importancia que ten a investigación, a innovación e o desenrolo tecnolóxico no avance socioeconómico e cultural da sociedade  |

| Resultados da aprendizaxe  |     |                                     |     |
|--|-----|-------------------------------------|-----|
| Resultados de aprendizaxe  |     | Competencias / Resultados do título |     |
| O alumno será capaz de seleccionar, instalar, configurar e xestionar o software básico para o procesamento de datos masivos. | AP1 | BP2                                 | CP1 |
|  | AP2 | BP6<br>BP8<br>BP10                  |     |
| O alumno será capaz de implementar códigos nalgunha linguaxe especializada no procesamento de datos masivos.                 | AP2 | BP1                                 | CP1 |
|  |     | BP2                                 |     |
|  |     | BP10                                |     |



|   |            |            |            |
|---|------------|------------|------------|
| O alumno coñecerá e aprenderá a utilizar algunhas das ferramentas dispoñibles para Data Engineering (en particular, par Inxesta/Almacenamento/Procesado/Visualización). | AP1<br>AP2 | BP1<br>BP2 | CP1<br>CP4 |
| O alumno adquirirá a habilidade necesaria para a procura, selección e manexo de recursos (bibliografía, software, etc.) relacionados con Big Data.                      | AP1<br>AP2 | BP1<br>BP6 | CP1<br>CP4 |

| Contidos                             |  |
|--------------------------------------|--|
| Temas                                | Subtemas   |
| 1. Introducción a Data Engineering   | 1.1 HPC vs Big Data: similitudes e diferencias no tratamento de datos<br>1.2 Tecnoloxías Hardware e Software para High Performance Data Engineering<br>1.3 Data Engineering en infraestructuras HPC vs entornos Cloud  |
| 2. Introducción a Analítica de Datos | 2.1 Exploratory Data Analytics<br>2.2 Introducción a Machine Learning  |
| 3. Etapas de Data Engineering        | 3.1 Modelado (Formatos, Compresión, Deseño de Esquemas)<br>3.2 Inxesta (Periodicidade, Transformaciones, Ferramentas)<br>3.3 Almacenamento (HDFS y BBDD NoSQL, HBase, MongoDB, Cassandra)<br>3.4 Procesado (Batch, Real-Time)<br>3.5 Orquestación<br>3.6 Análise (SQL, Machine Learning, Graphs, UI)<br>3.7 Gobernanza<br>3.8 Integración con BI (Visualización) |
| 4. Casos de Uso                      | 4.1 Aplicaciones en Internet das Cousas (entornos Smart e Industria 4.0)<br>4.2 Aplicaciones en ciencias e enxeñarías  |

| Planificación            |                           |   |                         |              |
|--------------------------|---------------------------|---|-------------------------|--------------|
| Metodoloxías / probas    | Competencias / Resultados | Horas lectivas (presenciais e virtuais) | Horas traballo autónomo | Horas totais |
| Sesión maxistral         | A1 A2 B1 C4               | 18                                      | 0                       | 18           |
| Prácticas de laboratorio | B1 B8 B10                 | 20                                      | 60                      | 80           |
| Traballos tutelados      | A1 A2 B1 B2 B8            | 0                                       | 45                      | 45           |
| Discusión dirixida       | B6 C1 C4                  | 4                                       | 2                       | 6            |
| Atención personalizada   |                           | 1                                       | 0                       | 1            |

\*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

| Metodoloxías             |   |
|--------------------------|---|
| Metodoloxías             | Descrición  |
| Sesión maxistral         | Impartidas polos profesores da asignatura. Inclúen exposición de material teórico, así como de seminarios.                        |
| Prácticas de laboratorio | Resolución de problemas e casos prácticos.  |
| Traballos tutelados      | Realización de prácticas de maior entidade de forma semiautónoma, guiados polos profesores da asignatura.                         |
| Discusión dirixida       | Orientación para a realización dos traballos individuais ou en grupo, resolución de dúbidas e actividades de avaliación continua. |

| Atención personalizada  |  |
|---|--|
| Metodoloxías  | Descrición   |
| Discusión dirixida<br>Prácticas de laboratorio<br>Traballos tutelados | Durante as prácticas de laboratorio, traballos tutelados, e discusións dirixidas, os estudantes poderán presentar preguntas, dúbidas, etc. O profesor, atendendo ás súas solicitudes, repasará conceptos, resolverá novos problemas ou utilizará calquera actividade que considere adecuada para resolver as cuestións expostas. |



## Avaliación

| Metodoloxías             | Competencias / Resultados | Descrición  | Cualificación |
|--------------------------|---------------------------|---|---------------|
| Prácticas de laboratorio | B1 B8 B10                 | Evaluación de las prácticas llevadas a cabo por los estudiantes.        | 50            |
| Traballos tutelados      | A1 A2 B1 B2 B8            | Evaluación de los trabajos tutelados desarrollados por los estudiantes. | 50            |

## Observacións avaliación

|  |
|--|
| <p>Non presentado: Considerarase non presentado @ alumn@ que non entregue ningunha práctica nin traballo academicamente dirixido.</p> <p>Segunda oportunidade (extraordinaria - xuño/xullo): Volver a realizar aquelas prácticas e traballos tutelados que non se entregaran ou versións melloradas dos xa entregados.</p> <p>Para os casos de realización fraudulenta de exercicios ou probas será de aplicación o recollido na Normativa de avaliación do rendemento académico dos estudantes e de revisión de cualificacións.</p> <p>En concreto, a realización fraudulenta das probas ou actividades de avaliación, unha vez comprobada, implicará directamente a cualificación de suspenso na convocatoria en que se cometa: o/a estudante será cualificado con ?suspenso? (nota numérica 0) na convocatoria correspondente do curso académico, tanto se a comisión da falta se produce na primeira oportunidade como na segunda. Para isto, procederase a modificar a súa cualificación na acta de primeira oportunidade, se fose necesario.</p> |
|--|

## Fontes de información

|                                    |   |
|------------------------------------|---|
| <b>Bibliografía básica</b>         | <ul style="list-style-type: none"> <li>- Tom White (2015). Hadoop: The Definitive Guide. O'Reilly (4ª ed.)</li> <li>- Wes McKinney (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly (2ª ed.)</li> </ul> |
| <b>Bibliografía complementaria</b> | - Alex Holmes (2014). Hadoop in practice. Manning (2ª ed.)  |

## Recomendacións

### Materias que se recomenda ter cursado previamente

### Materias que se recomenda cursar simultaneamente

### Materias que continúan o temario

## Observacións

|  |
|--|
| <p>Recomendacións para o estudo da materia Debido ao forte compoñente práctico é recomendable ir facendo as actividades prácticas e traballos academicamente dirixidos de forma regular ao longo do cuadrimestre. O coñecemento do inglés tanto falado como escrito é imprescindible dado que a bibliografía e as conferencias externas poden desenvolverse en inglés. Observacións Farase un uso intensivo de ferramentas de comunicación online: videoconferencia, chat, etc. As sesións presenciais serán gravadas para ou revisión posterior. Ademais, farase uso da ferramenta Aula CESGA para a distribución de contidos, creación de foros de discusión, etc... As ferramentas software utilizadas nesta materia son xeralmente open-source ou teñen licenza gratuita para estudantes. Perspectiva de xénero-Segundo se recolle nas distintas normativas de aplicación para a docencia universitaria incorporase a perspectiva de xénero nesta materia (usarase linguaxe non sexista, utilizarase bibliografía de autores/as de ambos sexos, propiciarase a intervención en clase de alumnos e alumnas...)-Traballarase para identificar e modificar prexuizos e actitudes sexistas e influirase na contorna para modificalos e fomentar valores de respecto e igualdade.-Deberanse detectar situacións de discriminación por razón de xénero e proponerse accións e medidas para corrixilas.</p> |
|--|



(\*A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías