UNIVERSIDAD
COMPLUTENSE
MADRID

nticmaster

Máster de Formación Permanente en

# BIG DATA & DATA ENGINEERING

3ª EDICIÓN

Facultad de Estudios Estadísticos
Universidad Complutense de Madrid

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

# PROGRAMA

Módulo · Python para Desarrolladores

Módulo · Programación en Scala

Módulo · Arquitecturas de datos

Módulo · Modelado de datos

Módulo · Bases de datos NoSQL

Módulo · Apache Kafka y procesamiento en tiempo real

Módulo · Apache Spark

Módulo · Ingestas y Lagos de datos

Módulo · Pipelines de datos en cloud

Módulo · Arquitecturas basadas en contenedores

Módulo · Machine Learning y Deep Learning

Módulo · Productivizar un modelo

TFM · Trabajo Final de Máster

# La importancia del Data Engineering

## NECESIDAD DE GESTIONAR DATOS

Las empresas y organismos están adoptando rápidamente la **transformación digital** para aprovechar el valor de los **datos masivos** en la toma de decisiones. El Big Data se ha convertido en un recurso valioso para la gestión empresarial, y lo que comenzó como una ventaja competitiva se ha vuelto esencial para mantenerse relevante. El dato es considerado como el petróleo del siglo XXI, y aquellos que puedan aprovechar su potencial estarán en una posición privilegiada para **innovar y liderar** en sus respectivos mercados.

## ECOSISTEMA BIG DATA

El impacto de la información a gran escala va más allá del ámbito matemático o estadístico, ya que tiene aplicaciones prácticas en diversos campos empresariales, gubernamentales, científicos y sociales. El Big Data se ha convertido en un recurso fundamental para **afrontar situaciones complejas en tiempo real**, permitiendo tomar decisiones informadas y generar conocimiento valioso en áreas como la medicina, la seguridad, el marketing, entre otros. El Big Data es una herramienta imprescindible en el mundo actual que ayuda a resolver problemas y aprovechar oportunidades en diversos entornos y sectores.

## INNOVACIÓN IMPULSADA POR LA INGENIERÍA DE DATOS

Los Data Engineers no solo contribuyen a la eficiencia operativa, sino que también juegan un papel clave en impulsar la innovación. Su capacidad para gestionar y organizar grandes volúmenes de datos permite **descubrir patrones ocultos y nuevas oportunidades de negocio**, lo que se traduce en productos y servicios más competitivos y alineados con las demandas del mercado actual.

## EL ROL DEL DATA ENGINEER

Las metodologías de la Ciencia de Datos para extraer valor de los datos, basadas en Machine Learning e Inteligencia Artificial, requieren que, primero, se hayan definido infraestructuras y arquitecturas que les permitan acceder a estos datos. Los Ingenieros de Datos son los encargados de **crear flujos de datos con los que proveer a los Científicos de Datos** que los analizarán, y definir procesos y estándares para desplegar en producción dichos modelos predictivos y explotar sus resultados, cerrando el ciclo de aporte de valor en la empresa. Para ello utilizan herramientas especializadas en datos masivos y tecnologías cloud.

## OPTIMIZACIÓN EN LA EXPERIENCIA DEL CLIENTE

La labor de los Data Engineers permite a las organizaciones optimizar la experiencia del cliente de manera más precisa. A través de la estructuración y procesamiento eficiente de grandes volúmenes de datos, los ingenieros de datos pueden **identificar patrones de comportamiento y preferencias**, lo que facilita la creación de modelos y sistemas que mejoren productos y servicios. Esto ayuda a alinear mejor la oferta con las expectativas del consumidor, aumentando tanto la satisfacción como la lealtad.

# ¿Por qué estudiar un Máster de Data Engineering?

Los **ingenieros de datos son la base para construir equipos de analítica avanzada**. Crean la infraestructura para que los científicos de datos trabajen, configuran sus entornos, evangelizan a los equipos de analítica y a los analistas de BI para que sigan las buenas prácticas en el uso de los datos, y definen y monitorizan los flujos adecuados para suministrarles datos curados y de calidad a todos los demás usuarios de una organización.

Esto incluye el uso de herramientas cloud (orquestar flujos y bases de datos diversas) y también la implementación de procesos específicos de ingestas y tratamientos, puesto que los datos en crudo habitualmente distan aún de ser útiles para perfiles analíticos.

Asimismo, los ingenieros de datos, en conjunción con los equipos de analítica, determinan **las prácticas de ingeniería que deben seguirse para desplegar y llevar a producción los modelos analíticos, y se encargan de su mantenimiento y monitorización operativa** - esto es, que permanezcan vivos y funcionales - al mismo tiempo que los científicos de datos monitorizan las métricas que se van obteniendo.

## Alta Demanda Laboral

El campo de los datos está en constante expansión. Las empresas de todos los sectores buscan profesionales capaces de gestionar, analizar y transformar grandes volúmenes de datos en información valiosa para la toma de decisiones. Especializarte en Big Data y Data Engineering te posicionará en un mercado con alta demanda y numerosas oportunidades.

## Actualización Constante

El campo del Data Engineering está en continua evolución con nuevas herramientas, tecnologías y enfoques. Un máster te permitirá mantenerte actualizado en las últimas novedades y aprender a adaptarte rápidamente a las tendencias emergentes en la industria de los datos.

## Oportunidades Profesionales

Las salidas profesionales en Big Data y Data Engineering son muy diversas, abarcando roles como analista de datos, ingeniero de datos, arquitecto de datos y científico de datos, entre otros. Un máster te preparará para enfrentar estos desafíos y desarrollar habilidades técnicas avanzadas.

## Desarrollo de Proyectos Reales

Un máster en Data Engineering te ofrecerá la oportunidad de trabajar en proyectos reales, aplicando tus conocimientos a situaciones del mundo empresarial. Estas experiencias prácticas serán clave para fortalecer tus habilidades y mejorar tu empleabilidad.

## Networking

Estudiar un máster en Big Data y Data Engineering te brindará la oportunidad de conectar con profesionales del sector, compañeros de clase, profesores expertos y empresas. Este entorno te permitirá generar relaciones clave para tu desarrollo profesional y acceder a oportunidades laborales.

## Ventaja Competitiva

Tener un máster en Data Engineering te dará una ventaja competitiva en el mercado laboral, diferenciándote como un profesional preparado para abordar los complejos desafíos de la gestión y análisis de grandes volúmenes de datos.

## Mayor Potencial de Ingresos

Los profesionales en Big Data y Data Engineering suelen tener un alto potencial de ingresos debido a la demanda de personas con habilidades técnicas avanzadas. Las empresas están dispuestas a pagar más por aquellos que puedan transformar datos en información estratégica que impulse la toma de decisiones.

## Flexibilidad Laboral

El conocimiento en Big Data ofrece flexibilidad en el tipo de empresas para las que puedes trabajar, desde startups hasta grandes corporaciones. Además, muchos roles relacionados con los datos permiten trabajar de forma remota, lo que te da mayor control sobre tu vida profesional.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

## Duración

1 año académico

## Modalidades

Presencial, Semipresencial y Online

## Creditos ECTS

60

# INFORMACIÓN

## Modalidad Presencial

Viernes tarde y sábados mañana en la universidad

## Modalidad Semipresencial

3 Semanas en presencial en la universidad y en la plataforma online

## Modalidad Online

100% desde nuestra plataforma online

# ¿Por qué estudiar en la Universidad Complutense de Madrid?

La Universidad Complutense de Madrid (UCM) es una de las instituciones educativas más destacadas de Europa, reconocida por el prestigioso QS World Ranking como la mejor de España. Ofrece una amplia gama de oportunidades y beneficios para los estudiantes, así como una excelencia académica reconocida, una calidad docente de primer nivel. Ofrece alrededor de 90 títulos de grado y más de 30 dobles grados, más de 200 programas máster, además de estudios de formación permanente. La UCM tiene más de 500 años de historia y reconocimiento social. La Universidad Complutense de Madrid es la universidad española de referencia en 5 continentes.

El prestigio de la universidad está avalado por 7 Premios Nobel, 20 Príncipes de Asturias, 7 Premios Cervantes, Premios Nacionales de Investigación y a la Excelencia. La Universidad Complutense de Madrid tiene estudiantes de más de 90 países y convenios con universidades de los 5 continentes.

# ¿Por qué estudiar un Máster en Formación Permanente?

*Si hay algo que afianza los conceptos teóricos de un programa educativo es la práctica.*

Nuestros módulos formativos combinan una base teórica con ejercicios prácticos basados en situaciones reales de las empresas. Además, todos los módulos se evalúan con tareas prácticas, no con exámenes, tratándose de un programa de configuración eminentemente práctica.

La preparación del Trabajo Final de Máster (TFM) garantiza la puesta en práctica de todos los conceptos adquiridos a lo largo del curso, capacitando definitivamente al alumno para asumir responsabilidades dentro de un entorno laboral real.

# PROGRAMA

*Los módulos de aprendizaje de Data Engineering más completos enseñados de forma eficaz para el alumno.*

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# MÓDULO

# Python para Desarrolladores

Siendo el primer módulo del máster, se profundiza en la **utilización del lenguaje Python para desarrollo de software**. Python es el lenguaje que mayor auge ha experimentado en los últimos años, tanto en el área de la Ciencia de Datos como de la Ingeniería de Datos, por su versatilidad y sencillez. De hecho, se utilizará en varios de los módulos posteriores, dado que dispone de paquetes específicos para integrarse con distintas tecnologías.

Dado que el alumno ya tiene experiencia programando en algún lenguaje, aquí **se repasarán los conceptos aplicados a Python y se profundizará en aspectos concretos** que debe conocer bien un Ingeniero de Datos en su día a día, ligados a los estándares del Desarrollo de Software.

## Índice de contenidos:

- Compiladores, intérpretes, lenguajes interpretados y compilados.
- Flujo de control, estructuras de datos básicas (listas, diccionarios).
- Manejo de DataFrames de Pandas.
- Control de excepciones:
  - Expresiones regulares.
- Programación Orientada a Objetos.
- Versionado de código con Git.
- Desarrollo de un paquete de Python.
- Tests unitarios y paquetes específicos.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

MÓDULO

# Programación en Scala

El lenguaje Scala se ha impuesto como uno de los estándares para la **creación de flujos de datos, motores de ingesta y preparación del dato**.

Se trata de un lenguaje con características avanzadas que combina el paradigma orientado a objetos, habitual en la industria del desarrollo de software desde hace años, con el paradigma funcional, algo novedoso y que da lugar a códigos elegantes, concisos y muy expresivos.

Se introducirá al alumno en los **conceptos tradicionales de la programación orientada a objetos y funcional en el lenguaje Scala**, así como a las construcciones habituales de este lenguaje y sus características funcionales fundamentales, apoyándonos en el IDE más popular entre los ingenieros de datos.

## Índice de contenidos:

- Introducción a Scala. Relación con Java y la JVM.
- Construcciones para flujo de control propias de Scala.
- Programación Orientada a Objetos en Scala. Clases, traits, objects. Similitudes y diferencias con Java.
- Pattern matching y características funcionales avanzadas.
- Introducción al desarrollo en IntelliJ.
- Creación de proyectos basados en sbt.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# MÓDULO

# Arquitecturas de datos

Hoy en día se están generando más datos que nunca y, para seguir siendo competitivas, las empresas deben recopilarlos, almacenarlos, procesarlos y analizarlos de manera eficaz. Ahí es donde entran en juego las arquitecturas de datos.

En este módulo, se introducirán y explicarán los **componentes clave de las arquitecturas de datos modernas**, incluyendo las fuentes de datos, la capa de ingesta, la capa de almacenamiento, la capa de procesamiento, capa de servicio y la capa de presentación. Se presentará su evolución a lo largo de los años, los principales patrones arquitectónicos actuales y su implementación en los distintos proveedores cloud. El alumno podrá descubrir cómo estas arquitecturas **permiten a las organizaciones manejar y procesar grandes volúmenes de datos de manera eficiente**, proporcionando información en tiempo real sobre sus operaciones.

Al final de este curso, el alumno tendrá conocimiento general para diseñar e implementar arquitecturas de datos modernas que permitan a cualquier empresa, gestionar y analizar sus datos de manera efectiva y tomar mejores decisiones basadas en datos.

## Índice de contenidos:



UNIVERSIDAD COMPLUTENSE MADRID

ntic master

# MÓDULO

# Modelado de datos

Las bases de datos relacionales, aquellas en las que la **información está almacenada en forma de tabla**, siguen siendo las más frecuentes en gran cantidad de empresas. Por razones históricas estas empresas almacenan en ellas los datos de su negocio.

También en muchas de las tecnologías más recientes del ecosistema Big Data existe la posibilidad de **definir tablas y modelos de datos que se basan en el lenguaje SQL**. Por esto, un ingeniero de datos frecuentemente debe llevar a cabo migraciones y definiciones o redefiniciones de modelos de datos basados en tablas, para lo cual necesita conocimientos clásicos de vv.

Durante los procesos de desarrollo de flujos de datos también es necesario consultar frecuentemente bases de datos utilizando el lenguaje de consulta SQL, el más extendido en la actualidad.

En la primera parte del módulo se revisará el **lenguaje SQL**, desde las consultas más básicas a las más avanzadas. A continuación, se presentarán los **principios de diseño habituales para definir modelos de datos**. Esto incluye conceptos como maestros, modelos en estrella, copo de nieve, claves primarias y externas, restricciones, índices, etc.

## Índice de contenidos:

- Concepto de modelado de datos.
- El lenguaje de consulta SQL:
  - Sentencias DDL.
  - Clave primaria, externa y restricciones
  - Sentencias DML.
  - Ejemplos de queries con SELECT, FROM, WHERE, GROUP BY, HAVING, agregaciones, subquerie, etc.
- Diseño de modelado de datos:
  - Modelo Entidad-Relación.
  - Modelo relacional.
  - Normalización de bases de datos.
  - Modelos de datos: en estrella, en copo de nieve, concepto de maestros.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

MÓDULO

# Bases de Datos NoSQL

El módulo de Bases de Datos NoSQL está diseñado para **proporcionar un conocimiento práctico de las bases de datos NoSQL**. El programa aborda conceptos fundamentales como la distribución de datos, la escalabilidad, la consistencia eventual y las operaciones CRUD, utilizando diferentes tipos de bases de datos NoSQL como documentales, clave-valor, columna y grafo. También cubre las bases de datos más populares en cada categoría, incluyendo MongoDB, Cassandra, Redis y Neo4j. Con una combinación de teoría, ejemplos prácticos y proyectos, los estudiantes aprenderán cómo implementar soluciones de bases de datos NoSQL de manera efectiva.

Las bases de datos NoSQL se están convirtiendo en una opción cada vez más popular para las organizaciones debido a su **capacidad para manejar grandes cantidades de datos y su flexibilidad para adaptarse a los cambios de datos**, lo que las hace ideales para big data y aplicaciones en tiempo real. Al adquirir estas habilidades, como ingeniero de datos contarás con conocimientos para afrontar los desafios del futuro en la gestión de datos no estructurados. En una era en la que los datos son el nuevo petróleo, aprender sobre bases de datos NoSQL puede ser una pieza clave para abrir nuevas oportunidades en una amplia gama de industrias, desde la tecnología hasta la salud, las finanzas y más allá.

## Índice de contenidos



UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

MÓDULO

# Apache Kafka y procesamiento en tiempo real

Apache Kafka es una plataforma de paso de mensajes en near real time, empleada para **comunicar instantáneamente distintas aplicaciones conectadas a él**, al estilo de un gran bus de datos común por el que circula la información de una empresa para ser procesada y utilizada por distintos departamentos de maneras muy diferentes. Es capaz de procesar billones de mensajes al día en near real time. Podríamos definir su arquitectura como un **sistema de logs distribuido**. Kafka sólo tiene una tarea: escribir mensajes en forma de log cumpliendo siempre sus dos grandes razones de ser: rapidez y fiabilidad en la entrega y proceso de mensajes

Sobre esta idea, se desarrolla un **ecosistema en forma de APIs y clientes caracterizado por su ligereza y sencillez** que permiten el escalado dinámico y elástico de la aplicación, que no depende necesariamente del escalado del propio cluster de máquinas en las que se está ejecutando Kafka. Estos son los motivos por los que el ecosistema Kafka se ha convertido en el estándar de facto para numerosos casos de uso alrededor de las arquitecturas de dato que van desde el paso de mensajes en tiempo real para comunicar sistemas heterogéneos, hasta el "sistema nervioso central" del dato de una compañía, que **permite que todos los departamentos vean siempre dato actual y habilita estrategias de gobierno de datos**. Incluso se aplica para la comunicación entre microservicios en aplicaciones descentralizadas. En definitiva, Kafka facilita la implementación de cualquier solución orientada a Eventos (EDA o Event Driven Architecture).

En este curso revisaremos la **Arquitectura Kafka y su ecosistema**, profundizando en temas como el escalado o las distintas estrategias de despliegue, pero sobre todo en sus distintas **APIs y abstracciones** (desde consumer producer a KSQLDB pasando por Kafka Connect), el d**esarrollo de aplicaciones clientes** para el streaming de datos en los lenguajes más usados (Python, Java/Scala) y el **diseño de pipelines** de datos eficientes y sencillas en (near) tiempo real.

## Índice de contenidos

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

# MÓDULO

# Apache Spark

Este módulo introduce las tecnologías Big Data y su motivación en el contexto actual de la era digital y las necesidades de las empresas. Proporciona a los estudiantes una **comprensión profunda de cómo funcionan estos sistemas de procesamiento de datos distribuidos y cómo aprovecharlos** para procesar grandes cantidades de datos de manera eficiente y efectiva.

El curso se centrará en Apache Spark, sin duda **la tecnología más demandada para procesamiento de grandes volúmenes de datos**, que constituye el día a día de los equipos de ingenieros de datos de todo el mundo. Describiremos su filosofía basada en un grafo de ejecución (DAG) y sus implicaciones.

A continuación, el alumno profundizará en el estudio de cada uno de los módulos, en especial Spark SQL, MLlib y Structured Streaming. Se desplegará un cluster de Spark en la plataforma de Databricks sobre Azure, actualmente una de las combinaciones más extendidas en la empresa privada, y sobre él se mostrará la aplicación de cada uno de los conceptos.

## Índice de contenidos

- Introducción a las tecnologías Big Data.
- Apache Spark:
  - Arquitectura de Spark.
- Módulos de Spark:
  - Spark SQL.
  - Spark MLlib.
  - Structured Streaming.
- Grafos con el paquete GraphFrames.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

MÓDULO

# Ingestas y Lagos de datos

Prepárate para sumergirte y bucear en el apasionante mundo de los lagos de datos. En el mundo actual, las empresas necesitan como parte de su arquitectura de datos, un componente que les permita **almacenar, procesar y analizar grandes cantidades de datos**. Estos componentes son los data lakes (o lagos de datos), y su última evolución, los lakehouses.

Los data lakes son **repositorios centralizados para almacenar cualquier tipo de dato sin requerir una estructura** previa. Los **lakehouses** son una evolución de los lagos de datos que combinan la escalabilidad y flexibilidad de los datalakes con el rendimiento y la fiabilidad de los data warehouses tradicionales. Sin embargo, administrarlos y operarlos de forma óptima es un desafío que requiere planificación y diseño para garantizar una organización adecuada de los datos en el mismo.

Al realizar este curso, tendrás los conocimientos necesarios para **diseñar, implementar, organizar y administrar estos componentes**. Además, aprenderás sobre formatos de almacenamiento, incluyendo el manejo de formatos de fichero específicos para lagos y la funcionalidad que ofrecen; estudiaremos los patrones de ingesta y lo relativo a la promoción de datos entre las distintas capas lógicas

## Índice de Contenidos

- Formatos de almacenamiento.
- Patrones de ingesta en batch y en tiempo real.
- Capas lógicas de un lago.
- Construcción de lagos de datos con herramientas cloud.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

MÓDULO

# Pipelines de datos en cloud

El curso se centra en el **diseño y desarrollo de soluciones de procesamiento de datos en la plataforma Microsoft Azure**. Exploraremos en detalle las herramientas específicas de Azure empleadas por los ingenieros de datos, tanto en la fase de ingesta como en las etapas posteriores. El enfoque principal será lograr un aprendizaje significativo a través de **proyectos personales guiados por el tutor de manera directa** y evaluación basada en la adquisición de competencias.

Los estudiantes aprenderán los **fundamentos y las mejores prácticas para construir pipelines de datos eficientes y escalables**, que permitan la ingesta, transformación, almacenamiento y análisis de grandes volúmenes de datos.

El curso cubre conceptos esenciales como el almacenamiento y procesamiento de datos en la nube, la arquitectura de pipelines de datos y las herramientas disponibles en el proveedor Microsoft Azure para su implementación. También se hará hincapié en la seguridad y monitorización de los pipelines, así como en la optimización del rendimiento y la escalabilidad de las soluciones implementadas.

El alumno se familiarizará con los servicios de Azure para cada etapa del ciclo de vida de los datos: ingesta, orquestación y coordinación de flujos de datos desde diversas fuentes mediante Azure Data Factory. Además, podrán gestionar la ingesta de datos en tiempo real y procesar eventos a gran escala utilizando Azure Event Hubs. Se profundiza en Storage Account Gen2 para administrar y proteger grandes volúmenes de datos. Explorarán el análisis de datos a gran escala, creando flujos de trabajo colaborativos y analizando datos masivos con Azure Databricks.

Podrán utilizar **diferentes lenguajes de programación** o emplear **Azure Synapse Analytics** como herramienta integral para el análisis y procesamiento de datos de alto rendimiento.

Tendrán asimismo la oportunidad de **trabajar con diversas bases de datos relacionales y no relacionales**, como Azure SQL Database y Azure Cosmos DB, así como otros servicios transversales relevantes.

## Índice de Contenidos

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

MÓDULO

# Arquitecturas basadas en contenedores

El curso proporcionará a los participantes los conocimientos y habilidades necesarios para **comprender, diseñar, desarrollar y desplegar arquitecturas basadas en microservicios utilizando API REST**, así como una introducción a las herramientas **Docker** y **Kubernetes**. Las arquitecturas basadas en microservicios son una estrategia que permite desarrollar sistemas escalables, flexibles y fáciles de mantener. En este curso, los participantes aprenderán los principios de las arquitecturas basadas en microservicios y API REST, las mejores prácticas para su diseño, desarrollo y despliegue, y cómo utilizar Docker y Kubernetes para gestionar eficientemente los microservicios.

Los objetivos del módulo incluyen:

- Comprender los conceptos fundamentales de las arquitecturas basadas en microservicios.
- Diseñar y desarrollar una arquitectura basada en microservicios de API Rest utilizando buenas prácticas.
- Utilizar Docker para crear y gestionar contenedores de microservicios.
- Utilizar Kubernetes para desplegar, escalar y gestionar microservicios en entornos de producción.
- Implementar estrategias de comunicación y gestión de datos entre microservicios.
- Aplicar prácticas recomendadas para garantizar la seguridad y la fiabilidad en una arquitectura de microservicios.
- Realizar la monitorización y el mantenimiento de una arquitectura basada en microservicios.

## Índice de contenidos:

1. **Introducción a las arquitecturas basadas en microservicios y API Rest:**
   - ¿Qué son las arquitecturas basadas en microservicios? ¿Qué es un API REST?
   - Principios y características de las arquitecturas basadas en microservicios de API Rest. Comparación con otros enfoques arquitectónicos.
   - Estudio de casos y ejemplos prácticos.
2. **Diseño y desarrollo de microservicios de API Rest:**
   - Principios de diseño para microservicios de API Rest.
   - Descomposición de aplicaciones monolíticas en microservicios de API Rest.
   - Tecnologías y herramientas para el desarrollo de microservicios de API Rest.
   - Gestión de la comunicación y la interacción entre microservicios de API Rest.
   - Implementación y despliegue de microservicios de API Rest utilizando Docker.
3. **Despliegue y gestión de microservicios de API Rest con Kubernetes:**
   - Introducción a Kubernetes y la orquestación de contenedores.
   - Despliegue de microservicios en clústeres de Kubernetes.
   - Escalado y gestión de microservicios utilizando Kubernetes.
   - Configuración y gestión de la comunicación entre microservicios en Kubernetes.
4. **Seguridad, monitorización y mantenimiento de microservicios de API Rest:**
   - Consideraciones de seguridad para las arquitecturas basadas en microservicios de API Rest.
   - Implementación de autenticación y autorización de microservicios de API Rest.
   - Monitorización y registro de microservicios en Kubernetes.
   - Estrategias de escalado y gestión de la carga en Kubernetes.
   - Pruebas, depuración y mantenimiento de microservicios.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

# MÓDULO

# Machine Learning y Deep Learning

Uno de los objetivos de la ingeniería de datos es **dar soporte a la creación de modelos de aprendizaje automático** que extraigan valor a los datos de una empresa y ayuden al negocio. Esto ocurre tanto antes de que se diseñe un modelo predictivo, suministrando los datos adecuados, como en el momento de ponerlo en producción.

Por ello, es necesario que un ingeniero de datos esté familiarizado con las **técnicas que aplican los científicos de datos**, con el fin de comprender las necesidades de estos, lo cual favorece la sinergia entre equipos de ambos perfiles y acelera la entrega de valor.

En este módulo, se presentan los **fundamentos del Machine Learning**, las **técnicas principales** que lo componen en el ámbito del aprendizaje supervisado, no supervisado y por refuerzo, así como las **fortalezas, limitaciones y métricas** necesarias para evaluar el funcionamiento de cada modelo.

El módulo se plantea desde un punto de vista eminentemente **práctico**, con una orientación específica a lo que el ingeniero de datos necesita entender.

Se complementa con una **introducción al Deep Learning**, el conjunto de técnicas basadas en redes neuronales que actualmente constituyen una verdadera tendencia, en especial en lo que respecta a procesamiento de lenguaje natural con LLM (Large Language Models) y redes generativas de contenidos de tipo textual e imagen.

## Índice de contenidos:

- Introducción al Machine Learning.
- Aprendizaje supervisado con Python.
- Aprendizaje no supervisado con Python.
- Utilización del paquete scikit-learn.
- Evaluación de un modelo entrenado.
- Redes neuronales: forward y backpropagation.
- Autoencoders. Transformers. Redes generativas. Ejemplos utilizando Keras.
- Large Language Models (LLMs) actuales.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# MÓDULO

# Productivizar un modelo

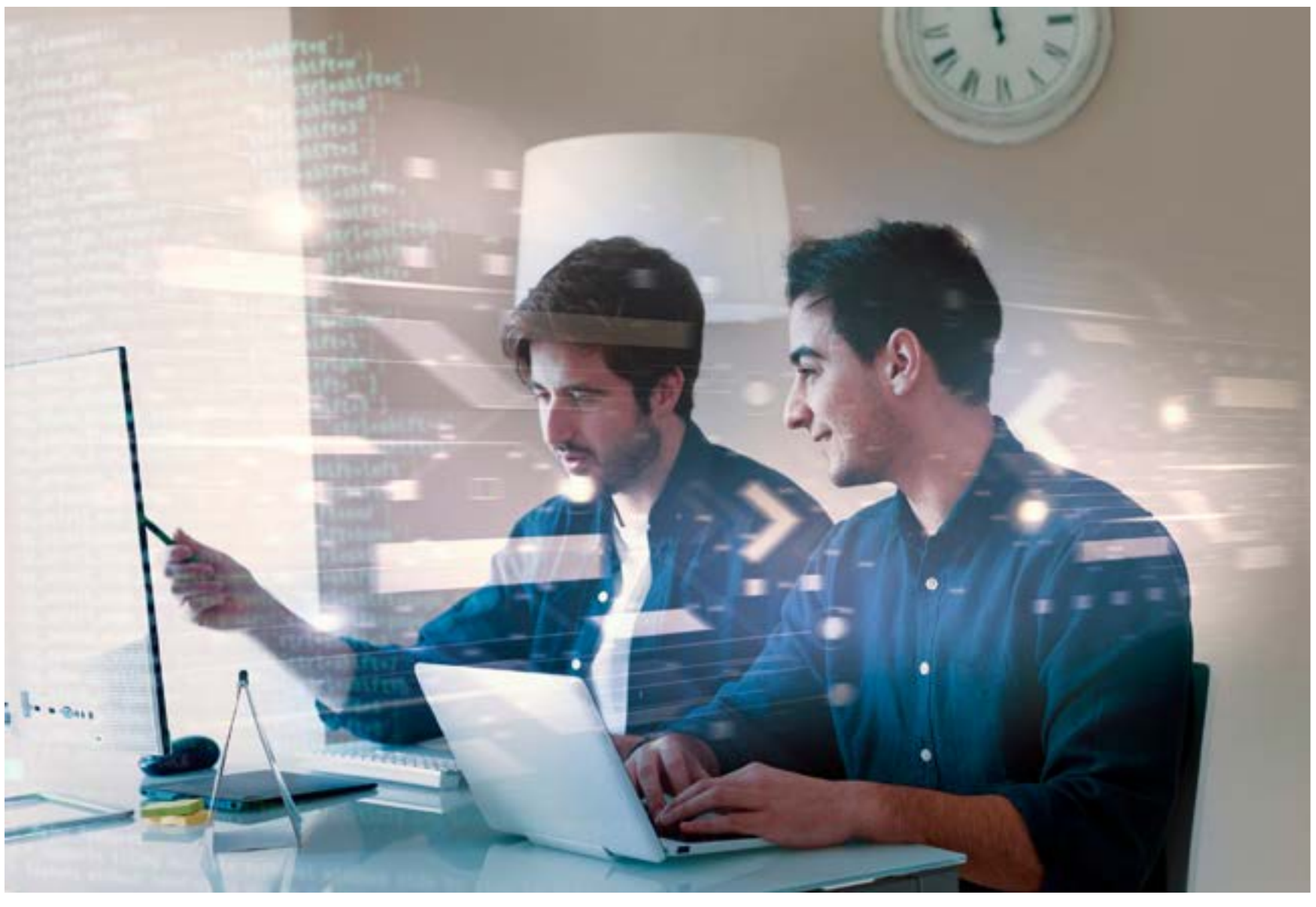
La productivización se refiere al proceso de **llevar los modelos de Inteligencia Artificial (IA) y sus resultados a un entorno de producción para que puedan ser utilizados de manera efectiva y generar valor en el mundo empresarial**. Los modelos de IA tienen un ciclo de vida que nace en su fase de desarrollo, pero es en la implementación, versionado y mantenimiento donde necesita convivir dentro de una infraestructura tecnológica.

Desde la recolección de datos y entrenamiento inicial, hasta la evaluación y monitorización, linaje, iteración y eventual reentrenamiento para mantener su rendimiento óptimo a lo largo del tiempo existen herramientas y técnicas que nos permiten cumplir con directrices de buenas prácticas que requiere el mercado.

En este curso estudiaremos la automatización y estandarización de todo el ciclo de vida de modelos de IA siguiendo estándares de **MLOps**. Al aplicar MLOps, los equipos de Data Science y desarrollo pueden colaborar de manera efectiva para acelerar la implementación de modelos en producción, facilitar la auditoría y garantizar la reproducibilidad de los resultados

## Índice de contenidos:

- Ciclo de vida de modelos AI y MLOps.
- Arquitecturas de implementación y despliegue en infraestructuras escalables.
- Versionado de modelos, linaje y trazabilidad.
- Monitorización y métricas de rendimiento.
- Reentrenamiento automático y despliegue gradual de modelos.



TFM

# Trabajo Fin de Máster

Asimilados todos los conceptos previos, llega el momento de poner a prueba todos los conocimientos adquiridos en el máster.

El alumno planteará una **estrategia global de inteligencia de datos para una empresa**, basándose en diferentes técnicas y softwares de apoyo de entre los existentes en el mercado.

El trabajo de fin de máster es una **parte crucial del programa**, ya que permite a los estudiantes aplicar todos los conocimientos adquiridos en el curso en un proyecto práctico y relevante en el mundo real.

El trabajo de fin de máster proporciona una oportunidad para que los estudiantes demuestren su capacidad para **analizar, procesar y utilizar datos de manera efectiva para resolver problemas complejos y tomar decisiones informadas**. También les permite desarrollar habilidades de presentación y comunicación al presentar sus hallazgos y resultados a una audiencia de expertos.

Además, el trabajo de fin de máster puede ser una oportunidad para que los estudiantes trabajen en **colaboración con empresas u organizaciones**, lo que les permite obtener experiencia práctica en un entorno profesional y crear conexiones valiosas para su carrera.

# EQUIPO DOCENTE

*Clases y tutorías con grandes profesionales del sector del Big Data y Data Engineering.*

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# EQUIPO DOCENTE

## Directivos

### Cristóbal Pareja Flores

Director. Catedrático EU en la UCM. Con más de 30 años como docente, Cristóbal es matemático especializado en Ciencias de la Computación, Doctor en Informática. Además, es Decano de la Facultad de Estudios Estadísticos y Vicedecano de Postgrado e Investigación.

### Jose Carlos Soto Gómez

Co-Director del Máster. Socio Fundador de NTIC Master y Aplimovil. Amplia experiencia en proyectos nacionales e internacionales en IT y analítica en empresas como Banco de España, NEC, Telefónica, Vodafone, Orange, medios de comunicación...

## Coordinadores

### David del Ser

**Coordinador Máster**

David es Lic. en Marketing por ESIC, Honours Degree in Business Administration por Humberside University, MBA por UNED, Máster Dirección Financiera, Máster Marketing Digital, Máster en Big Data. Especialista en el desarrollo de negocio y transformación digital en Ntic Master. Gran experiencia profesional trabajando en Grupo Iberostar, Grupo Avintia, entre otras.

### Cristóbal Martínez Martínez

**Coordinador Máster**

Cristóbal es Ingeniero informático. Director de IT en Aplimovil y Ntic Master. Profesor máster marketing digital de la UCM, UNED, Cámara de Comercio y CEEIC. Experto en sistemas y procesos informáticos. Gran experiencia profesional trabajando en empresas referentes como NEC, BNP Paribas, Banco de España, Vodafone.

# EQUIPO DOCENTE

## Docentes

### Cristóbal Pareja Flores

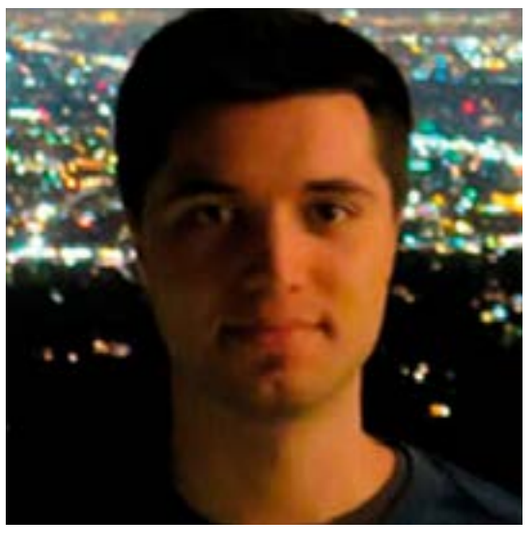
**Catedrático EU en la UCM**

Con más de 30 años como docente, Cristóbal es matemático especializado en Ciencias de la Computación, Doctor en Informática. Además, es Decano de la Facultad de Estudios Estadísticos y Vicedecano de Postgrado e Investigación.

### José Javier Galán Hernández

**Responsable de Sistemas**

José Javier es Ingeniero Informático y trabaja como Responsable de Sistemas CED. Además es profesor asociado UCM. Ha trabajado en proyectos de sistemas en El Corte Inglés y Comel, entre otros.



### Pablo J. Villacorta

**Data Scientist & ML Engineer en Seguros Santalucía**

Pablo es Doctor en Ciencias de la Computación e IA, Ingeniero Informático y Lic. en Estadística por la Univ. de Granada. Desarrollador certificado en Spark por Databricks, trabaja desde hace 8 años como Data Scientist / ML Engineer especializado en la creación y puesta en producción de modelos basados en Spark.

### Marlon Cárdenas Bonett

**Líder de Data Science en Sopra Steria**

Marlon es responsable de Data Science en Sopra Steria, liderando varias iniciativas de proyectos en el área de la analítica avanzada en sectores diversos. Además, es arquitecto de soluciones especializado en bases de datos.

### Alberto González

**Arquitecto Cloud en Minsait**

Alberto es Ingeniero en Informática experto en arquitectura de datos, y soluciones tanto cloud como on-premise. Certificado como Solutions Architect Expert y Data Engineer Associate en cloud de Microsoft Azure. Actualmente diseña e implementa arquitecturas de datos cloud para diversos sectores.

### David Alonso García

**Profesor en la UCM**

David es Doctor en Geografía e Historia. Experto en gestión y desarrollo de la innovación y emprendimiento. Ha sido director de Compluemprende-Universidad Complutense de Madrid.

# EQUIPO DOCENTE

## Docentes

### Jorge Centeno

**Head of Data en Inditext Logistics SA**

Jorge es ingeniero de software con más de 15 años de experiencia. Ha desarrollado arquitecturas de datos en múltiples industrias como redes sociales, viajes, sector público y seguros, como arquitecto de soluciones y liderando equipos de Ingenieros de datos.

### Elena del Carmen Gavilán García

**Docente e investigadora en la UCM**

Graduada en estadística aplicada y Doctora en análisis de datos, Data Science. Docente e investigadora en la Universidad Complutense de Madrid

### Álvaro Bravo Acosta

**Ingeniero Técnico Informático en Sistemas**

Experto en Tecnologías Big Data, BI y Analítica. Gran experiencia en consultoras como Minsait, Sopra Steria, Everis o NTT, para externos como ISBAN y BBVA. Actualmente, DevOps en equipo de producto Frameworks en Strato.

### Eduardo Fernández Carrión

**Data Scientist, PhD.**

Eduardo es ingeniero informático y doctor en métodos estadísticos matemáticos para el tratamiento computacional de la información. Data Scientist & ML Engineer en Santalucia seguros. Experto Data Sciencist habiendo desarrollado su carrera profesional en StratioBD, Visavet, Everis, etc.

### Armando Heras

**Chief Digital Officer**

Armando es Licenciado en Economía por CUNEF. Experto en finanzas, auditorías y financiación, especialmente en ecosistemas emprendedores a través de la inversión, creación de empresas y desarrollo de proyectos.

### Luis Piñon Ferrer

**Data Scientist & Full Stack Developer Entrepreneur**

Graduado en Ingeniería Informática y en Matemáticas Computacionales trabaja tanto como Data Architect en Capgemini desarrollando apliaciones innavadoras de CloudOps Open Source. Experiencia en soluciones en la nube y arquitecturas escalables.

# EQUIPO DOCENTE

## Docentes

### Alberto Ezpondaburu

**NLP Specialist**

Alberto es ingeniero de telecomunicaciones, Lic. en Matemáticas. Experto en NLP y en la aplicación de técnicas de deep learning e inteligencia artificial. Ha trabajado como Data Scientist en varias empresas y creado empresas del ámbito AI.


### Pedro Pablo Malagon Amor

**Principal Architect en Google**

Pedro Pablo es ingeniero de sistemas. Experto en Big Data, Data Science, Data Engineer. Tiene una gran experiencia profesional de más de 23 años en Microsoft como principal architect y actualmente en Google.

### Yolanda García Ruiz

**Docente en la UCM**

Licenciada en C.C. Matemáticas en la especialidad de C.C. de la Computación por la Universidad Complutense de Madrid desde el año 1995 y doctora en Informática. Hasta su incorporación al mundo académico ha desarrollado su carrera profesional en el área de la informática en diferentes compañías.

### Daniel Martín García

**Docente en la UCM**

Daniel es graduado en estadística y doctorado en Data sSience. Trabaja como docente en la UCM, donde también forma parte de grupos de investigación.

### Gabriel Marín Díaz

**PhD, Análisis de Datos**

Gabriel es Licenciado en Matemáticas y Doctor en Análisis de Datos, Data Science. Tiene una amplia experiencia en el sector empresarial siendo consultor data Science, actividad que compatibiliza actualmente con la de profesor en la UCM

### Charles Flores Espinoza

**Big Data Engineer en Mercedes-Benz AG**

Charles es Ingeniero informático y experto Data Engineer & Data Architect, Scala, ML. Gran experiencia en el sector en empresas como StratioBD, Oliver Wyman, Vass o Bayes.

# EQUIPO DOCENTE

## Docentes

### Pablo Arcadio Flores Vidal

**Docente en la UCM**

Pablo A. Flores es licenciado en Estadística y doctor en Análisis y Ciencia de Datos. Actualmente es profesor en la Universidad Complutense de Madrid y desempeña el cargo de Delegado del Decano para Erasmus y Movilidad.

### Luis Fernando Llana Díaz

**Docente en la UCM**

Luis Fernando es doctor por la UCM. Docente en la UCM y forma parte de grupos de investigación: Design and Testing of Reliable Systems. Experto en Computer Science.

### Olga Marroquín Alonso

**Docente en la UCM**

Olga es licenciada y doctora en mátemáticas por la UCM. Docente en la UCM y forma parte de grupos de investigación. Experta en Data Science.

### Tú

**Futuro experto en Big Data y Data Engineer**

Porque con nosotros aprenderás Big Data, Data Engineering, Programación orientada al dato, Arquitectura de datos, etc. Pero al final el camino tienes que recorrerlo tú y quizás muy pronto estés aquí como profesor nuestro.

# SALIDAS PROFESIONALES

*El sector de Tecnologías de la Información continuará siendo el motor principal del crecimiento laboral, encabezando el top de empleos más demandados y mejor remunerados por las empresas.*

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# Salidas Profesionales

El Big Data y el Data Engineering ofrece una amplia variedad de salidas profesionales debido a su creciente relevancia en la estrategia empresarial moderna. Estas son solo algunas de las muchas oportunidades profesionales que podrás explorar dentro de este ámbito:v

*Con el avance de la tecnología y la evolución constante del entorno digital, es probable que sigan surgiendo nuevas especialidades y roles.*

**Gestor y auditor de infraestructuras para Big Data**

Responsable de planificar, gestionar y auditar los sistemas que permiten almacenar y procesar grandes volúmenes de datos de manera eficiente.

**Ingeniero de aprendizaje automático**

Desarrolla y optimiza algoritmos de machine learning para que los sistemas puedan aprender y mejorar automáticamente a partir de datos.

**Especialista en inteligencia de datos y de BI**

Se enfoca en analizar datos y generar informes para ayudar en la toma de decisiones estratégicas basadas en la inteligencia empresarial.

**Chief Data Officer**

Ejecutivo responsable de la estrategia de datos de una organización, asegurando su correcta gestión y uso para impulsar los objetivos de negocio.

**Data Analyst**

Analiza conjuntos de datos grandes y complejos para extraer conclusiones útiles que apoyen la toma de decisiones en las empresas.

**Data Consultant**

Proporciona asesoramiento experto en la gestión, análisis y uso estratégico de los datos para mejorar el rendimiento empresarial.

**Data Scientist**

Utiliza estadísticas, machine learning y otras herramientas avanzadas para interpretar datos complejos y ofrecer soluciones a problemas empresariales.

**Data Mining**

Especialista en técnicas de minería de datos que busca patrones, tendencias y relaciones ocultas en grandes bases de datos para extraer información útil.

**Arquitecto de datos**

Diseña y gestiona la infraestructura que permite almacenar y procesar grandes volúmenes de datos de manera eficiente. Elige las tecnologías adecuadas y garantiza que los datos sean accesibles, seguros y optimizados para su uso.

# ADMISIONES

*Tanto la preinscripción como la pre matrícula quedan abiertas hasta comenzar el curso académico o completar plazas.*

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# Proceso de admisión

### Preinscripción

Envío de solicitud para evaluar candidaturas.

### Evaluación

Evaluación de candidaturas.

### Pre-admisión

Confirmación como alumno del candidato.

### Admisión

Confirmación de plaza y formalización de la matrícula.

1. Preinscribirse cumplimentando el formulario ubicado en la pestaña “Preinscripción” de la web **www.masterdataengineeringucm.com**
2. Enviar la documentación requerida a fin de evaluar la candidatura.
3. Entrevista con el solicitante.
4. Confirmación de selección.
5. Realización de un pago inicial.

Tanto la preinscripción como la pre matrícula quedan abiertas hasta comenzar el curso académico o completar plazas, estableciéndose lista de espera si procede. Los alumnos deberán ingresar 600 euros en concepto de pago inicial para el Máster Presencial y 500 euros en concepto de pago inicial para el Máster Semipresencial y el Máster Online, los cuales les serán descontados del importe total de la matrícula. En ningún caso se tendrá derecho a devolución de dicha cantidad, a excepción de que no se llegara a celebrar el curso.

# Documentación requerida

## Alumnos con titulación de **España**

Los documentos identificativos requeridos para la inscripción en el Máster son:

- Fotocopia del documento de identidad/pasaporte.
- Certificado de notas oficial.
- Título universitario o resguardo de solicitud de título.
- Currículum Vitae.

## Alumnos con titulación de **Unión Europea**

- Currículum Vitae.
- Pasaporte/NIE (no válidas las cédulas de identificación de fuera de España).
- Título universitario (no es valido el certificado del título).
- Certificado oficial de notas.

*La documentación debe estar traducida al castellano por un traductor jurado homologado. (Solicitar listado oficial)

## Alumnos con titulación de **Fuera de la Unión Europea**

- Currículum Vitae.
- Pasaporte/NIE (no válidas las cédulas de identificación de fuera de España).
- Título universitario legalizado con la Apostilla de la Haya (no es valido el certificado del título).
- Certificado oficial de notas.

*La documentación debe estar traducida al castellano por un traductor jurado homologado. (Solicitar listado oficial)

# MODALIDADES

*La evaluación de los alumnos se realizará a lo largo de todo el programa a través de ejercicios y casos prácticos.*

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# Máster Presencial

### Duración

**1 Año / 520 horas**

**60 ECTS**

Inicio: Febrero de 2026
Fin: Febrero de 2027

Viernes: de 16:00 a 21:00 h
Sábados: de 09:00 a 14:00 h

### Lugar

**Facultad de Estudios Estadísticos**

Ciudad Universitaria Avenida Puerta de Hierro, s/n, 28040 Madrid

### Precio

**6.750 €**

**+ 40 € de tasas de secretaría**

Pregunta por nuestras becas, facilidades de pago, prácticas en empresas y bolsa de trabajo.

*Una vez finalizados y superados estos estudios, la Universidad Complutense de Madrid emitirá el título, conforme a las normas de admisión y matriculación de los títulos de Formación Permanente de la UCM*

## Metodología Presencial

El curso se impartirá en aulas de la Universidad Complutense de Madrid, en la Facultad de Estudios Estadísticos los viernes y sábados con masterclasses impartidas por diferentes expertos. La formación se realizará de forma tutorizada por los profesores. También se utilizará una plataforma de formación virtual para la comunicación entre los alumnos y profesores, creando una comunidad virtual de trabajo. Los distintos profesores de cada módulo, guiarán a los alumnos proponiendo actividades adicionales dependiendo del temario que se esté cubriendo en cada momento.

## Características plataforma On-line

La plataforma actuará como vía de comunicación entre el alumno y el entorno global de formación.

El estudiante tendrá información actualizada sobre los conceptos que se estén estudiando en cada momento, como enlaces a contenidos adicionales incluyendo noticias, artículos, etc.

Los alumnos deberán realizar y aprobar todas las prácticas de los distintos módulos, y realizar el trabajo fin de máster para poder aprobar el máster.

**La plataforma cuenta con:**

- Mensajería individualizada para cada alumno.
- Vídeos de las clases y de casos prácticos.
- Tutorías online con el profesorado.
- Documentación, noticas y contenidos.
- Foro de los módulos del máster.
- Comunicación con los profesores vía mensajería.
- Chat entre alumnos.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster

# Máster Semipresencial

## Duración

**1 Año / 520 horas**

**60 ECTS**

Inicio: Octubre de 2026
Fin: Octubre de 2027

Jueves y viernes: de 16:00 a 21:00 h
Sábados: de 09:00 a 14:00 h

## Lugar

**Online con presencialidad de 3 semanas**

**Facultad de Estudios Estadísticos**

Ciudad Universitaria Avenida Puerta de Hierro, s/n, 28040 Madrid

## Precio

**5.700 €**

**+ 40 € de tasas de secretaría**

Pregunta por nuestras becas, facilidades de pago, prácticas en empresas y bolsa de trabajo.

*Una vez finalizados y superados estos estudios, la Universidad Complutense de Madrid emitirá el título, conforme a las normas de admisión y matriculación de los títulos de Formación Permanente de la UCM*

## Metodología Semipresencial

La formación se realizará de forma tutorizada por los profesores. Se utilizará una plataforma de formación virtual para la comunicación entre los alumnos y profesores, creando una comunidad virtual de trabajo. Los distintos profesores de cada módulo, guiarán a los alumnos proponiendo actividades adicionales dependiendo del temario que se esté cubriendo en cada momento. La modalidad semipresencial contempla la realización de 3 semanas presenciales con masterclasses impartidas por diferentes expertos para preparar los TFM y hacer networking.

## Características plataforma On-line

La plataforma actuará como vía de comunicación entre el alumno y el entorno global de formación.

El estudiante tendrá información actualizada sobre los conceptos que se estén estudiando en cada momento, como enlaces a contenidos adicionales incluyendo noticias, artículos, etc.

Los alumnos deberán realizar y aprobar todas las prácticas de los distintos módulos, y realizar el trabajo fin de máster para poder aprobar el máster.

**La plataforma cuenta con:**

- Mensajería individualizada para cada alumno.
- Vídeos de las clases y de casos prácticos.
- Tutorías online con el profesorado.
- Documentación, noticas y contenidos.
- Foro de los módulos del máster.
- Comunicación con los profesores vía mensajería.
- Chat entre alumnos.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

# Máster Online

### Duración

**1 Año / 520 horas**

**60 ECTS**

Inicio: Febrero de 2026
Fin: Febrero de 2027

### Lugar

**Plataforma Online**

### Precio

**4.650€**

**+ 40€ de tasas de secretaría**

Pregunta por nuestras becas, facilidades de pago, prácticas en empresas y bolsa de trabajo.

*Una vez finalizados y superados estos estudios, la Universidad Complutense de Madrid emitirá el título, conforme a las normas de admisión y matriculación de los títulos de Formación Permanente de la UCM*

## Metodología 100% Online

La formación se realizará de forma tutorizada por los profesores. Se utilizará una plataforma de formación virtual para la comunicación entre los alumnos y profesores, creando una comunidad virtual de trabajo. Los distintos profesores de cada módulo, guiarán a los alumnos proponiendo actividades adicionales dependiendo del temario que se esté cubriendo en cada momento.

## Características plataforma On-line

La plataforma actuará como vía de comunicación entre el alumno y el entorno global de formación.

El estudiante tendrá información actualizada sobre los conceptos que se estén estudiando en cada momento, como enlaces a contenidos adicionales incluyendo noticias, artículos, etc.

Los alumnos deberán realizar y aprobar todas las prácticas de los distintos módulos, y realizar el trabajo fin de máster para poder aprobar el máster.

**La plataforma cuenta con:**

- Mensajería individualizada para cada alumno.
- Vídeos de las clases y de casos prácticos.
- Tutorías online con el profesorado.
- Documentación, noticas y contenidos.
- Foro de los módulos del máster.
- Comunicación con los profesores vía mensajería.
- Chat entre alumnos.

# Contacto

**Teléfono de información**

+34 687 30 04 04

**Teléfono de admisiones**

+34 667 89 05 83

**Correo electrónico**

info@masterdataengineeringucm.com

**Sitio Web**

www.masterdataengineeringucm.com

*La dirección del máster se reserva el derecho de modificar, suprimir y actualizar los profesores, la información y el programa del máster.

UNIVERSIDAD COMPLUTENSE MADRID

nticmaster