UNIVERSIDAD COMPLUTENSE MADRID

ntic master



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

1. Python para desarrolladores
2. Arquitecturas de datos
3. Desarrollo asistido por agentes
4. Modelado de Data Warehouses
5. Modelado de datos no estructurados
6. Kafka y procesamiento en tiempo real
7. Apache Spark para pipelines de datos
8. Cloud Data Engineering en Azure
9. Diseño de ingestas y Data Lakehouses
10. MLOps
11. Arquitecturas de agentes
12. Diseño y evaluación de agentes
13. Arquitecturas basadas en contenedores
14. Trabajo Final de Máster

UNIVERSIDAD COMPLUTENSE MADRID | nticmaster

La importancia del Data Engineering

Escalabilidad del almacenamiento

Los volúmenes de datos crecen de forma exponencial. Contar con arquitecturas escalables permite almacenar y consultar grandes cantidades de información sin degradar el rendimiento ni disparar los costes de infraestructura.

Calidad e integridad de los datos

Unos datos limpios y consistentes son la base de cualquier decisión fiable. Los procesos de validación, limpieza y control de calidad evitan que errores en el origen se propaguen por todo el pipeline.

Procesamiento en tiempo real

Cada vez más negocios necesitan reaccionar al instante. Tecnologías de streaming como Kafka o Spark Structured Streaming permiten analizar eventos según se producen, habilitando decisiones inmediatas.

Gobernanza y trazabilidad

Saber de dónde viene cada dato, quién lo ha transformado y cómo se ha usado es clave para cumplir normativas y generar confianza en los análisis y modelos construidos sobre esa información.

Arquitecturas cloud y distribuidas

Servicios como Azure Data Lake, Databricks o Snowflake permiten separar almacenamiento y cómputo, escalando cada recurso de forma independiente y pagando solo por lo que realmente se utiliza.

Automatización de pipelines (DataOps)

Orquestar la ingesta, transformación y carga de datos con herramientas como Airflow o dbt reduce errores manuales, acelera los despliegues y facilita la reproducibilidad de los procesos analíticos.

Interoperabilidad y estándares abiertos

Formatos como Parquet, Delta Lake o Iceberg permiten que distintos motores de procesamiento trabajen sobre el mismo dato sin duplicarlo, evitando el bloqueo a un único proveedor tecnológico.

¿Por qué estudiar un Máster de Ingeniería de Datos?

En un mundo donde el volumen de datos crece sin parar, contar con una formación especializada en Big Data e Ingeniería de Datos es una ventaja competitiva clave. Este programa ofrece una formación práctica en las tecnologías más innovadoras para el tratamiento de datos, incluyendo arquitecturas distribuidas, procesamiento en tiempo real y diseño de pipelines, permitiendo a los estudiantes aplicar soluciones de datos en múltiples sectores.

El máster está diseñado con un enfoque práctico, donde los alumnos trabajan en proyectos reales que les permiten construir sistemas de datos de manera eficiente y aplicable al entorno laboral. Desde la ingesta y modelado de datos hasta el despliegue de infraestructuras en la nube, los estudiantes explorarán las prácticas más demandadas en la actualidad.

En una era impulsada por la digitalización, este máster capacita a los profesionales con conocimientos actualizados sobre ingeniería de datos y el uso de herramientas avanzadas para diseñar, construir y mantener sistemas de datos de forma eficaz.

Además, los alumnos tendrán la oportunidad de interactuar con expertos del sector, analizar casos de éxito en la gestión de proyectos de datos y desarrollar su propia visión estratégica sobre la arquitectura de datos moderna. El objetivo fundamental del máster es formar profesionales capaces de diseñar e implementar soluciones de datos en su campo de especialización, optimizando el aprovechamiento de la información y desarrollando soluciones innovadoras para el futuro digital.

¿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.

El prestigio de la universidad está avalado por:

QS STARS ★★★★★

La universidad hispanohablante con **la máxima calificación en el ranking** de la reconocida consultora **QS World University Rankings**.

7 Premios Nobel, 20 Príncipes de Asturias, 7 Premios Cervantes, etc.

Una de las **instituciones educativas más destacadas** de España y Europa.

Más de **500 años de historia** y reconocimiento social.

Más de 90 títulos de grado, 30 dobles grados, 200 programas máster, y **estudios de formación permanente.**

Excelencia académica reconocida, una calidad docente de primer nivel.

Estudiantes de **más de 90 países** y convenios con universidades de los 5 continentes.

¿Por qué estudiar un Máster en Formación Permanente?

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.

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



UNIVERSIDAD COMPLUTENSE MADRID | nticmaster

INFORMACIÓN

Duración	**1 año académico**
Creditos ECTS	**60**

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

PROGRAMA

Los módulos más completos de Big Data & Data Engineering, diseñados para un aprendizaje eficaz y accesible para el alumno.



UNIVERSIDAD COMPLUTENSE MADRID

ntic master

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

MÓDULO

Python para desarrolladores

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, lo utilizarás en varios de los módulos posteriores, dado que dispone de paquetes específicos para integrarse con distintas tecnologías.

Como ya tienes experiencia programando en algún lenguaje, repasa los conceptos aplicados a Python: los entornos de desarrollo, los intérpretes, la gestión de dependencias y los entornos virtuales, además de las funciones y las estructuras de datos básicas.

Trabaja con las estructuras de datos propias de la ciencia de datos en NumPy, Pandas y Polars, y profundiza en los aspectos concretos que un Ingeniero de Datos debe conocer bien en su día a día: las expresiones regulares, la programación orientada a objetos y el control de excepciones.

Cierra el módulo con la parte más ligada a los estándares del Desarrollo de Software: organiza tus proyectos, **desarrolla tus propios paquetes de Python y asegura su calidad con tests unitarios**.

Contenidos:

- ➔ Entornos de desarrollo, intérpretes, dependencias y entornos virtuales.
- ➔ Funciones y estructuras de datos básicas.
- ➔ Estructuras de datos para ciencia de datos: NumPy, Pandas y Polars.
- ➔ Expresiones regulares.
- ➔ Programación Orientada a Objetos.
- ➔ Control de excepciones.
- ➔ Proyectos y desarrollo de paquetes Python.
- ➔ Tests unitarios.

MÓDULO

Arquitecturas de datos

Descubre por qué entran en juego las 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.

Conoce 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, la capa de servicio y la capa de presentación. Recorre su evolución a lo largo de los años, los principales patrones arquitectónicos actuales y su implementación en los distintos proveedores cloud.

Comprende 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**.

Termina el módulo con el 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.

Contenidos:

- ➔ Introducción a las arquitecturas de datos.
- ➔ Fuentes de datos actuales.
- ➔ Capa de ingesta.
- ➔ Capa de almacenamiento.
- ➔ Capa de procesamiento.
- ➔ Capa de servicio.
- ➔ Capa de presentación.
- ➔ Patrones arquitectónicos de arquitecturas de datos.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

MÓDULO

Desarrollo asistido por agentes

Aprende a trabajar de forma **asistida por agentes de IA para multiplicar tu productividad** sin perder el control sobre el software que produces. La Inteligencia Artificial generativa está transformando la forma en que los ingenieros de datos escriben código, diseñan pipelines y automatizan tareas.

Estudia las **principales herramientas del ecosistema, como Claude Code, GitHub Copilot y los asistentes integrados en el IDE y en el terminal**, explorando sus diferencias, sus casos de uso y cómo integrarlas de forma efectiva en un flujo de trabajo de ingeniería de datos profesional. Gestiona el contexto, las instrucciones, MCP, skills, hooks y subagentes, y revisa, testea, valida y supervisa el código generado por IA.

Cierra el módulo con el desarrollo basado en especificaciones, un enfoque para definir de forma explícita requisitos, restricciones y criterios de aceptación antes de delegar tareas al agente, facilitando **resultados más controlables, revisables y alineados con las necesidades del proyecto**.

Contenidos:

- ➔ Agentes de código y uso de herramientas.
- ➔ Integración de agentes en el flujo de desarrollo e ingeniería de datos.
- ➔ Gestión de contexto, instrucciones, MCP, skills, hooks y subagentes.
- ➔ Revisión, testing, validación y supervisión del código generado por IA.
- ➔ Introducción al desarrollo basado en especificaciones.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

MÓDULO

Modelado de Data Warehouses

Organiza adecuadamente los datos, el primer paso para después explotarlos de manera más sencilla. Entiende que el modelado de datos es mucho más que crear tablas y relaciones en una base de datos: es **diseñar una representación abstracta de la información que capture su significado y sus interconexiones**, el puente entre los requisitos del negocio y la implementación técnica.

Distingue las bases de datos de los Data Warehouses y trabaja el **modelado conceptual con diagramas** Entidad-Relación y el **lógico con el modelo relacional y la normalización**. Una buena práctica de modelado garantiza la integridad, consistencia y accesibilidad de la información, elementos cruciales para la toma de decisiones informadas y la creación de sistemas robustos.

Explora el **modelo dimensional de Kimball**, base del Data Warehousing, con sus tablas de hechos y tablas de dimensiones y los modelos en estrella y en copo de nieve.

Termina el módulo **aplicando todos estos conocimientos en casos prácticos de modelado**.

Contenidos:

- Bases de Datos y Data Warehouses.
- Concepto de modelado de datos.
- Diseño de modelos de datos:
  - Modelo Entidad-Relación.
  - Modelo relacional.
  - Normalización de bases de datos.
  - Modelado dimensional: tablas de hechos y tablas de dimensiones.
  - Modelo de datos en estrella.
  - Modelo de datos en copo de nieve.

MÓDULO

Modelado de datos no estructurados

Utiliza y modela los sistemas de gestión de bases de datos NoSQL, **esenciales en las arquitecturas de Big Data**. Repasa un poco de historia de las bases de datos relacionales y compara ambos enfoques: el modelo entidad-relación frente al de agregación, y ACID frente a BASE.

Distingue los **principales tipos de bases de datos NoSQL** (clave-valor, documentales, XML, grafos y Column Family) y sus casos de uso, cuándo conviene utilizarlas y qué implica el teorema CAP.

Trabaja con **MongoDB y sus principales operaciones**: CRUD, dominar el find y proyectar los campos en los resultados de búsqueda. Compárala después con Cassandra.

Cierra el módulo aprendiendo a **modelar, implementar y optimizar soluciones NoSQL para grandes volúmenes de datos**, con el foco en la escalabilidad, la flexibilidad y el rendimiento en el procesamiento y almacenamiento de datos no estructurados y semiestructurados.

Contenidos:

- ➔ Introducción a las bases de datos NoSQL.
- ➔ Un poco de historia sobre las BD relacionales.
- ➔ Modelo de datos (entidad relación vs agregación).
- ➔ BD Relacionales VS NoSQL (ACID vs BASE).
- ➔ Diferentes tipos de BD NoSQL (key-value, documentales, XML, Grafos, Column Family).
- ➔ Cuándo utilizar NoSQL.
- ➔ Teorema del CAP.
- ➔ Mongo VS Cassandra.

MÓDULO

Kafka y procesamiento en tiempo real

Comunica **instantáneamente distintas aplicaciones con Apache Kafka**, una plataforma de paso de mensajes en near real time que funciona como un gran bus de datos común por el que circula la información de una empresa para que la procesen y utilicen distintos departamentos. Es capaz de **procesar billones de mensajes al día**.

Entiende su arquitectura como un sistema de logs distribuido, con dos razones de ser (rapidez y fiabilidad) y un ecosistema de APIs y clientes ligeros que permiten escalar la aplicación sin depender del escalado del propio clúster. Descubre **por qué se ha convertido en el estándar de facto de las arquitecturas orientadas a eventos** (EDA): del paso de mensajes entre sistemas heterogéneos y microservicios al "sistema nervioso central" del dato de una compañía.

Trabaja con KRaft, que sustituye a Zookeeper, la Admin API, **el escalado y las estrategias de despliegue, y sobre todo con sus APIs y abstraccione**s: Producer/Consumer y sus semánticas de entrega, Kafka Connect, Kstreams y KSQL, desarrollando aplicaciones cliente en Python y Java/Scala.

Termina el módulo diseñando pipelines de datos eficientes y sencillos pensados para el (near) tiempo real.

Contenidos:

- Arquitectura y Conceptos Básicos:
  - KRAFT (bye bye Zookeeper).
  - Kafka Admin API.
  - El LOG distribuido y sus conceptos básicos.
- Producer/Consumer API:
  - Conceptos básicos Producer.
  - Conceptos básicos Kafka Consumer.
  - Console Producer/Consumer.
  - Semánticas de entrega.
  - Java/Scala/Python Producer/Consumer.
- Kafka Connect.
- Kstreams.
- KSQL.
- Pipelines de Datos pensadas en tiempo real.

MÓDULO

Apache Spark para pipelines de datos

Introdúcete en las tecnologías Big Data y su motivación en el contexto actual de la era digital y las necesidades de las empresas. Comprende **cómo funcionan los sistemas de procesamiento de datos distribuidos** y cómo aprovecharlos para procesar grandes cantidades de datos de manera eficiente y efectiva.

Céntrate en Apache Spark, sin duda l**a tecnología más demandada para el procesamiento de grandes volúmenes de datos** y el día a día de los equipos de ingenieros de datos de todo el mundo. Conoce su arquitectura y su filosofía basada en un grafo de ejecución (DAG), con sus implicaciones.

Profundiza en cada uno de sus módulos, en especial **Spark SQL, MLlib y Structured Streaming, y trabaja también con grafos** mediante el paquete GraphFrames.

Cierra el módulo **desplegando un clúster de Spark** en la plataforma Databricks sobre Azure, actualmente una de las combinaciones más extendidas en la empresa privada, y aplicando sobre él cada uno de los conceptos.

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

Cloud Data Engineering en Azure

Especialízate en el **diseño y desarrollo de soluciones de procesamiento de datos en Microsoft Azure**, mucho más allá de lo puramente académico. Con un enfoque 100% práctico, proyectos personales guiados por el tutor y una evaluación basada en competencias, aprende a **construir pipelines eficientes, escalables y sostenibles** y a gestionar el ciclo de vida completo del dato, de la ingesta a la explotación.

Domina los fundamentos de Azure: ingesta y orquestación con **Azure Data Factory**, eventos en tiempo real con **Azure Event Hubs**, almacenamiento seguro en **Storage Account Gen2** y bases de datos como **Azure SQL Database y Cosmos DB**. Entra a fondo en Microsoft Fabric con OneLake, Mirroring, Shortcuts, Dataflows, Pipelines y Direct Lake para explotar los datos desde Power BI sin duplicidades, preparando el terreno para los futuros Agentes de Datos.

Cubre **el ciclo de vida analítico con Databricks**: gobierno y linaje del dato con Unity Catalog, orquestación con Workflows y Lakeflow, SQL Warehouses y analítica en lenguaje natural con Genie AI/BI. Participa además en los debates técnicos de un ingeniero de datos senior: FinOps, gobierno y seguridad del dato, y soberanía frente al vendor lock-in con formatos abiertos.

Por último, desarrolla, monitoriza y optimiza un pipeline completo en un proyecto práctico transversal basado en un caso de uso real.

Contenidos:

- ➔ Introducción a la Realidad de los Datos: Fundamentos y mejores prácticas en arquitecturas cloud modernas.
- ➔ Diseño de Plataformas Sostenibles: Seguridad, gobierno del dato, FinOps (control de costes) y evasión del vendor lock-in mediante formatos abiertos.
- ➔ Ingesta y Orquestación Core: Integración de fuentes con Azure Data Factory / Microsoft Fabric y almacenamiento escalable en Storage Account Gen2 y OneLake.
- ➔ Streaming y Tiempo Real: Procesamiento masivo de eventos con Azure Event Hubs.
- ➔ Revolución Microsoft Fabric: Arquitecturas OneLake, Dataflows, Mirroring, Shortcuts y explotación eficiente con Direct Lake.
- ➔ Ecosistema Databricks Avanzado: Procesamiento a gran escala, orquestación (Workflows/Lakeflow), gobierno de extremo a extremo con Unity Catalog y GenAI aplicada a BI.
- ➔ Bases de Datos en Cloud: Integración con modelos relacionales y no relacionales (Azure SQL Database y Azure Cosmos DB).
- ➔ Proyecto Práctico Transversal: Desarrollo, monitorización y optimización de un pipeline completo basado en un caso de uso real.

MÓDULO

Diseño de ingestas y Data Lakehouses

Diseña Data Lakehouses, **la última evolución de los lagos de datos y el componente que las empresas necesitan hoy en su arquitectura para almacenar, procesar y analizar grandes cantidades de datos**. Entiende cómo estos repositorios centralizados, capaces de almacenar cualquier tipo de dato sin requerir una estructura previa, combinan la escalabilidad y flexibilidad de los data lakes con el rendimiento y la fiabilidad de los data warehouses tradicionales.

Aprende sobre **formatos de almacenamiento**, incluido el manejo de formatos de fichero específicos para lagos y la funcionalidad que ofrecen, y sobre los patrones de ingesta en batch y en tiempo real.

Organiza los datos en las **distintas capas lógicas de un Data Lakehouse** y gestiona su promoción entre ellas, porque **administrarlo y operarlo de forma óptima** es un desafío que requiere planificación y diseño.

Termina el módulo construyendo **Data Lakehouses con herramientas cloud**, con los conocimientos necesarios para diseñarlos, implementarlos, organizarlos y administrarlos.

Contenidos:

- ➔ Formatos de almacenamiento.
- ➔ Patrones de ingesta en batch y en tiempo real.
- ➔ Capas lógicas de un Data Lakehouse.
- ➔ Construcción de Data Lakehouses con herramientas cloud.

MÓDULO

MLOps

Da soporte a uno de los objetivos de la Ingeniería de Datos: la **puesta en producción y el mantenimiento en el tiempo de modelos de IA** que extraigan valor de los datos de una empresa y ayuden al negocio.

Descubre **cómo el Ingeniero de Datos participa allí donde el Científico de Datos no llega**: antes de que se elabore un modelo predictivo, diseñando el almacenamiento y los procesos de recolección de datos adecuados, y después, al desplegarlo en un entorno corporativo, acoplarlo al resto de sistemas de la compañía y monitorizarlo como software productivo.

Trabaja la **analítica predictiva, los tipos de problemas de Machine Learning y las métricas de rendimiento** de un modelo, y asimila las prácticas de MLOps y el ciclo de vida completo de un modelo, empezando por cómo convertir un modelo creado por un Data Scientist en un componente software.

Completa ese ciclo con el **versionado de modelos y agentes** con MLflow, la **orquestación y el despliegue** con Airflow y microservicios, la **monitorización y detección** de drift en producción, el **reentrenamiento y el despliegue** de modelos de agentes con MCP, siempre desde un enfoque eminentemente práctico.

Contenidos:

- ➔ Definición y tipos de problemas en Machine Learning.
- ➔ Métricas de performance de un modelo.
- ➔ Introducción a MLOps y ciclo de vida de un modelo.
- ➔ Monitorización de modelos y detección de drift en producción.
- ➔ Versionado de modelos y agentes con MLflow.
- ➔ Orquestación y despliegue con Airflow y microservicios.
- ➔ Despliegue de modelos de Agentes, y uso de MCPs.

MÓDULO

Arquitecturas de agentes

Estudia las bases de **los Large Language Models (LLMs) y su utilización para realizar tareas de forma autónoma y coordinada**, poniéndolas en práctica a la vez que construyes un agente autónomo.

En primer lugar, aprende **el uso de un LLM y su conexión con herramientas**, Function Calling y Model Context Protocol (MCP) para conectar modelos con infraestructuras de datos reales de forma estandarizada.

Pasa después a las **Arquitecturas de Razonamiento y Estado**, con patrones como ReAct o Chain-of-Thought aplicados a Data Engineering, y a una introducción a LangGraph para orquestar agentes.

Cierra el módulo con conceptos de LLMOps y **estrategias de observabilidad, trazas y seguridad** (guardarraíles) para validar la fiabilidad de los agentes.

Contenidos:

- ➔ Large Language Models (LLMs).
- ➔ Fundamentos de uso de un LLM y su conexión con herramientas.
- ➔ Arquitecturas de Razonamiento y Estado.
- ➔ LLMOps y estrategias de observabilidad, trazas, seguridad.

MÓDULO

Diseño y evaluación de agentes

Construye **agentes de IA robustos y evalúa su comportamiento** de forma sistemática, dos competencias críticas para cualquier ingeniero de datos que trabaje con sistemas basados en LLMs.

Empieza con el concepto de **RAG** (Retrieval-Augmented Generation): sus componentes, las opciones comerciales disponibles, **cómo programarlos** con LangChain y las métricas para evaluarlos.

Pasa al concepto de agente y aprende a definir su arquitectura (herramientas, memoria, planificación y orquestación) y a **combinar estos componentes para resolver problemas** complejos de data engineering de forma autónoma.

Termina el módulo con **métricas de calidad como faithfulness o relevance, técnicas de benchmarking, detección de alucinaciones y fallos, y estrategias de seguridad** y guardarraíles para garantizar la fiabilidad de los agentes en producción.

Contenidos:

- ➔ Retrieval-Augmented Generation (RAGs): componentes de un RAG y opciones comerciales disponibles.
- ➔ Programación de RAGs con LangChain.
- ➔ Concepto de agente.
- ➔ Arquitectura de un agente: herramientas, memoria y planificación.
- ➔ Métricas de evaluación de agentes (faithfulness, relevance, etc.).
- ➔ Benchmarking y detección de alucinaciones.
- ➔ Estrategias de seguridad y guardarraíles en producción.

MÓDULO

Arquitecturas basadas en contenedores

Comprende, **diseña, desarrolla y despliega arquitecturas basadas en microservicios** utilizando API REST, una estrategia que permite desarrollar sistemas escalables, flexibles y fáciles de mantener.

Aprende los principios de las arquitecturas basadas en microservicios y API REST y **las mejores prácticas para diseñar y desarrollar tus propios microservicios**.

Da tus primeros pasos con **Docker y Kubernetes** y utilízalos para desplegar y gestionar eficientemente los microservicios.

Cierra el módulo con la seguridad, la monitorización y el mantenimiento de los microservicios de API REST.

Contenidos:

- ➔ Introducción a las arquitecturas basadas en microservicios y API Rest.
- ➔ Diseño y desarrollo de microservicios de API Rest.
- ➔ Despliegue y gestión de microservicios de API Rest con Kubernetes.
- ➔ Seguridad, monitorización y mantenimiento de microservicios de API Rest.

MÓDULO

Trabajo Final de Máster: TFM en Data Engineering

El Trabajo Final de Máster (TFM) que realizarás al finalizar el máster se centrará en la ingeniería de datos. Este proyecto implicará el **desarrollo de una estrategia global de ingeniería de datos para una empresa**, basándose en diferentes técnicas y softwares de apoyo de entre los existentes en el mercado. También deberás detallar iniciativas y elaborar informes con conclusiones basadas en los resultados obtenidos.

El proceso de desarrollo del TFM **será guiado por las pautas proporcionadas por los profesores responsables, quienes estarán disponibles para resolver cualquier duda o consulta**.

Será una oportunidad clave para aplicar todo lo aprendido en los módulos previos y demostrar tu capacidad para desarrollar un proyecto de ingeniería de datos completo.

Además, contarás con la supervisión de los profesores y miembros de la dirección del máster, lo que te permitirá recibir orientación de expertos en el área.

EQUIPO DOCENTE

Clases y tutorías con grandes profesionales del sector de la Ingeniería de Datos.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

EQUIPO DOCENTE

Directivos



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 Carlos Soto Gómez

Co-Director. 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



Alberto González

Arquitecto Cloud en NFQ

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.



Daniel Martín

Profesor e Investigador Ayudante

Doctor por la Facultad de Estudios Estadísticos, UCM, Madrid.



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.



Jorge Centeno

Head of Data en Inditex 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.



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 aplicaciones innovadoras de CloudOps Open Source. Experiencia en soluciones en la nube y arquitecturas escalables.



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.

EQUIPO DOCENTE

Docentes



Pablo Hidalgo

Databricks Champion ML Engineer en Capgemini

Desarrollador de soluciones escalables para la conversión de texto a SQL mediante algoritmos de vanguardia, prompt engineering, LLM, interfaces de usuario y Databricks (Unity Catalog y SQL Warehouses).



Pablo J. Villacorta

Data scientist & ML Engineer en Seguros Santalucía

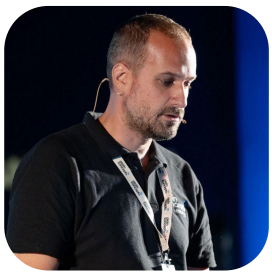
Pablo es Doctor en Ciencias de la Computación e IA, Ingeniero Informático y Lic. en Estadística por la Univ. de Granada. Desarrollador certificado en Spark por Databricks, trabaja desde hace 8 años como Data Scientist / ML Engineer especializado en la creación y puesta en producción de modelos basados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



Paul Van Branteghem

Asesor de Estrategia Nacional de IA en INECO (SEDIA) · Co-Founder de Spain AI

Licenciado en Física por la Universidade de Santiago de Compostela y Máster en Inteligencia Artificial Avanzada. Asesor de estrategia nacional de IA en INECO (SEDIA) y cofundador de Spain AI, una de las mayores comunidades de IA en España. Docente y experto en LLMs, RAG y AI Red-Teaming.



Manuel Rodrigo Cabello Malagón

Principal AI Research Engineer

Líder en estrategia e innovación en IA, con más de 12 años de experiencia en Inteligencia Artificial e ingeniería de software; actualmente dirige el área de investigación e innovación en IA en Plain Concepts.



Sergio Arpón

Arquitecto Big Data en Santalucía Seguros

Más de 15 años de experiencia en Business Analytics, Big Data y Machine Learning. Colabora en el desarrollo, supervisión y soporte de diferentes proyectos del ámbito de Arquitecturas Big Data en la Nube, realizando tareas tanto técnicas, como de gestión.



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.

EQUIPO DOCENTE

Docentes

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 nuestro profesor.

SALIDAS PROFESIONALES

La creciente digitalización y el crecimiento exponencial del volumen de datos han convertido a los expertos en Big Data e Ingeniería de Datos en perfiles indispensables para las empresas.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

UNIVERSIDAD COMPLUTENSE MADRID | ntic master

SALIDAS PROFESIONALES

El ecosistema de la ingeniería de datos demanda perfiles altamente especializados, capaces de diseñar, construir y explotar las infraestructuras que sostienen el valor de los datos en las organizaciones y empresas.

Ingeniero de datos (Data Engineer)

Diseña, construye y mantiene los pipelines que transforman datos brutos en información lista para su uso.

Ingeniero de Streaming / Real-Time

Gestiona la ingesta y el procesamiento de eventos en tiempo real con Kafka o Spark Streaming.

Arquitecto de Datos

Diseña la arquitectura global de datos de la organización, de la ingesta a la explotación.

Especialista en Calidad de Datos

Audita y asegura la fiabilidad de los datos a lo largo de todo el ciclo de vida del pipeline.

Ingeniero de Machine Learning (MLOps)

Despliega y mantiene modelos de IA en producción de forma fiable y escalable.

Analista de Datos (Data Analyst)

Evalúa el panorama de datos disponible para diseñar estrategias analíticas y de negocio.

Chief Data Officer (CDO)

Lidera la estrategia global de datos y la transformación digital en grandes organizaciones.

Ingeniero de IA y Agentes (AI Engineer)

Diseña y despliega agentes y sistemas basados en LLM conectados a los datos de la empresa.

Arquitecto de Datos Cloud

Diseña infraestructuras de datos y gestiona el despliegue en entornos AWS, Azure o Google Cloud.

Especialista en Gobernanza de Datos

Busca activamente inconsistencias y riesgos ocultos dentro de los sistemas de datos corporativos.

ADMISIONES

Tanto la preinscripción como la prematrícula quedan abiertas hasta comenzar el curso académico o completar plazas.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

PROCESO DE ADMISIÓN

Preinscripción
Envío de solicitud para evaluar candidaturas.

Preinscripción
Evaluación de candidaturas.

Preinscripción
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 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 prematrícula 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**

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

- 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**

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

- 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

Online, presencial o semipresencial. Elige la modalidad que mejor se adapta a ti.

UNIVERSIDAD COMPLUTENSE MADRID

ntic master

MÁSTER PRESENCIAL

Duración

1 Año / 520 horas

60 ECTS

Inicio: Septiembre de 2026
Fin: Septiembre de 2027

Viernes de 16h a 21h
Sábados de 9h a 14h

Lugar

Facultad de Estudios Estadísticos

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

Precio

6.900 €

+ 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, noticias y contenidos.
- ➔ Foro de los módulos del máster.
- ➔ Comunicación con los profesores vía mensajería.
- ➔ Chat entre alumnos.

MÁSTER SEMIPRESENCIAL

Duración

1 Año / 520 horas

60 ECTS

Inicio: Febrero de 2027
Fin: Febrero de 2028

3 semanas presenciales
Viernes de 16h a 21h
Sábados de 9h a 14h

Lugar

Online con presencialidad de 3 semanas

Facultad de Estudios Estadísticos

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

Precio

5.830 €

+ 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, noticias y contenidos.
- ➔ Foro de los módulos del máster.
- ➔ Comunicación con los profesores vía mensajería.
- ➔ Chat entre alumnos.

MÁSTER ONLINE

Duración

1 Año / 520 horas

60 ECTS

Inicio: Febrero de 2027
Fin: Febrero de 2028

Lugar

Plataforma Online

Precio

4.755€

+ 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, noticias 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

masterucm.com



UNIVERSIDAD COMPLUTENSE MADRID

ntic master