

AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial



ÍNDICE GENERAL DEL CURSO

Máster en ingeniería del dato



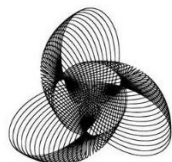
MÓDULO 1: Introducción a la Ingeniería de Datos y Ecosistema

- Panorama general: definición y objetivos de la ingeniería de datos
 - Arquitecturas tradicionales vs. modernas
 - Roles clave: Data Engineer, Data Scientist, Data Architect
 - Ciclo de vida del dato y tipos de proyectos
 - Herramientas esenciales y plataformas del ecosistema
-



MÓDULO 2: Python para Ingeniería de Datos

- Fundamentos de Python y buenas prácticas de código
- Gestión de entornos profesionales: virtualenv, pip, requirements.txt



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

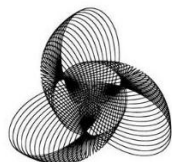
- Librerías esenciales: pandas, NumPy, requests, PyArrow
 - Lectura y escritura de archivos (CSV, JSON, Parquet, etc.)
 - Automatización de tareas y scripts ETL profesionales
-

✓ MÓDULO 3: SQL – Diseño y Consulta de Bases de Datos Relacionales

- Modelo relacional y diseño de bases de datos
 - DDL y DML: creación, inserción, modificación y borrado de datos
 - Consultas avanzadas: JOINS, subqueries, window functions
 - Índices, rendimiento y optimización de consultas
 - Administración básica y aspectos de seguridad en SQL
-

✓ MÓDULO 4: NoSQL – MongoDB, Cassandra y Redis

- Introducción y tipos de bases de datos NoSQL (Documento, Clave-Valor, Columnares)



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

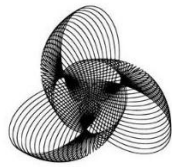
- MongoDB: Modelado, operaciones CRUD y consultas avanzadas
 - Cassandra: Fundamentos y modelado orientado a consultas
 - Redis: Estructura en memoria y caching de datos
 - Comparativa SQL vs NoSQL: cuándo usar cada tipo
-

✓ MÓDULO 5: Procesamiento y Limpieza de Datos

- Manejo de datos nulos, duplicados y errores comunes
 - Técnicas de normalización y estandarización de datos
 - Validación, profiling y control de calidad de datos
 - Transformaciones avanzadas con pandas y SQL
-

✓ MÓDULO 6: Ingesta de Datos desde Múltiples Fuentes

- Lectura de archivos planos (CSV, JSON, XML, Parquet)



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

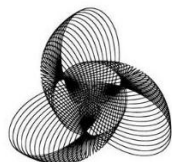
- Conexión y extracción desde bases de datos relacionales y NoSQL
 - Ingesta y procesamiento desde APIs REST y SOAP
 - Introducción al web scraping
 - Ingesta batch y en tiempo real (streaming)
-

✓ MÓDULO 7: Bash Scripting para Automatización

- Fundamentos de scripting en Bash
 - Manipulación avanzada de archivos y carpetas
 - Tuberías, redirección y comandos clave: grep, awk, sed
 - Automatización sencilla de procesos ETL
 - Programación y automatización de tareas con Crontab
-

✓ MÓDULO 8: Apache Airflow y Crontab

- Introducción a Apache Airflow: DAG, Tasks y Scheduler
- Instalación, configuración y despliegue de Airflow



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

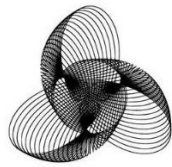
- Creación práctica de DAGs: PythonOperator y BashOperator
 - Gestión de dependencias, triggers y scheduling avanzado
 - Monitorización, logs y uso de la API Airflow
-

✓ MÓDULO 9: Contenedores y Virtualización – Docker y Kubernetes

- Conceptos clave de virtualización y contenedores
 - Docker: Creación de imágenes, contenedores y Docker Compose
 - Gestión de almacenamiento y redes en Docker
 - Introducción a Kubernetes: Pods, Servicios y despliegue básico
 - Aplicación de Docker y Kubernetes en pipelines de datos
-

✓ MÓDULO 10: Big Data – Ecosistema Hadoop y Spark

- Fundamentos del procesamiento distribuido y ecosistema Hadoop



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

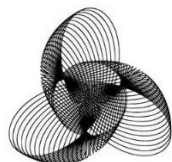
- HDFS: Arquitectura, estructura y operaciones básicas
 - Procesamiento distribuido con MapReduce
 - Apache Hive y Pig para consultas y análisis
 - Introducción a Apache Spark: arquitectura, conceptos y uso básico
-

✓ MÓDULO 11: PySpark y Databricks

- Instalación y configuración del entorno Databricks
 - Operaciones con DataFrames y RDDs en PySpark
 - Optimización y estrategias de particionamiento
 - Integración con otras fuentes (AWS S3, bases de datos)
-

✓ MÓDULO 12: Data Lake y Data Warehouse Moderno

- Diferencias y casos de uso: Data Lake vs Data Warehouse
- Modelos de datos modernos: Star Schema, Snowflake Schema, Data Vault



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

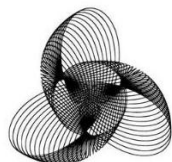
- Implementación práctica en AWS (S3, Glue, Redshift) y Snowflake
 - Buenas prácticas de gobierno y gestión de datos
-

✓ MÓDULO 13: Introducción a la Nube – AWS para Data Engineering

- Introducción a AWS y gestión de consola cloud
 - Uso avanzado de S3: almacenamiento, seguridad y versionado
 - Procesamiento con EC2 y Lambda
 - Pipelines y analítica con AWS Glue y Redshift
 - Gestión de permisos, seguridad e IAM
-

✓ MÓDULO 14: Data Streaming – Kafka y AWS Kinesis

- Fundamentos del procesamiento en streaming vs batch
- Arquitectura, componentes y uso de Apache Kafka
- Integración de Kafka con Spark y bases de datos
- Procesamiento y consumo en tiempo real con AWS Kinesis



AEDIA

Asociación Española para la Difusión de la Inteligencia Artificial

✓ MÓDULO 15: Introducción a la Observabilidad y Data Quality

- Conceptos esenciales de data quality y observabilidad
- Herramientas y plataformas: Great Expectations, OpenLineage
- Técnicas avanzadas de validación, alertado y gestión de incidentes
- Linaje de datos, trazabilidad y reporting avanzado

✓ MÓDULO 16: Proyecto Integrador y Buenas Prácticas Profesionales

- Planteamiento y desarrollo del proyecto integrador final
- Documentación técnica y presentación ejecutiva
- Exposición, defensa profesional y buenas prácticas de ingeniería de datos