

Aprender Online UTN

Diplomatura en Ingeniería de Datos

FACULTAD REGIONAL RECONQUISTA

APRENDERONLINE.FRRQ.UTN.EDU.AR

Acerca del curso

Programa integral enfocado en formar profesionales capaces de diseñar, construir y orquestar arquitecturas de datos modernas y escalables. Cubre desde la ingesta, automatización de pipelines y procesamiento masivo en entornos de Big Data, hasta la gobernanza y disponibilidad de datos para Business Intelligence y Machine Learning.

¿A quien va dirigido?

Está dirigido a desarrolladores, analistas de datos, administradores de bases de datos, profesionales de IT y estudiantes de carreras técnicas o afines que cuenten con conocimientos básicos de lógica de programación y manejo de datos, y deseen especializarse en el diseño y la gestión de infraestructura de datos de alto rendimiento.

Perfil Profesional

Al completar la diplomatura, la persona estará capacitada para:

- **Diseño de repositorios**: Modelar y administrar bases de datos SQL y NoSQL (BigQuery, MongoDB), así como Data Warehouses y Data Lakes.
- **Construcción de pipelines**: Diseñar, automatizar y orquestar flujos de datos (ETL/ELT) robustos utilizando Shell scripting, Apache Kafka y Apache Airflow.
- **Procesamiento de Big Data**: Procesar grandes volúmenes de información distribuida mediante motores como Apache Spark y Hadoop.
- **Gobernanza y seguridad**: Implementar estándares de calidad, seguridad, compliance y buenas prácticas a lo largo del ciclo de vida del dato.
- **Visualización de datos**: Crear paneles interactivos y comunicar hallazgos estratégicos utilizando Tableau y Looker.
- **Preparación para Machine Learning**: Estructurar y disponibilizar conjuntos de datos optimizados para el entrenamiento y despliegue de modelos predictivos.

Características generales

Modalidad

Online • On-demand • Asíncrona

Duración

20 semanas

Nivel de dificultad

Bajo

Dedicación

Moderada

Pre-requisitos

Conocimientos previos

Es recomendable contar con nociones básicas de lógica de programación (en cualquier lenguaje), familiaridad con el uso general de sistemas operativos y conceptos elementales de tablas o bases de datos (como Excel o SQL básico). Si bien el programa introduce las herramientas paso a paso, no está pensado para usuarios sin experiencia informática básica, sino para personas con curiosidad técnica y pensamiento analítico interesadas en adentrarse en la infraestructura de datos.

Requerimientos técnicos

- Computadora: Laptop o PC con Windows (64 bits), macOS o Linux (mínimo 8 GB de RAM).
- Permisos: Acceso de administrador para instalar programas y herramientas.
- Conectividad: Conexión a Internet estable.
- Cuentas: Navegador web actualizado y cuenta de correo electrónico activa.

Modulo 1: Introducción a la Ingeniería de Datos

- ¿Qué es la Ingeniería de Datos?
- El ecosistema de datos actual
- Evolución de la Ingeniería de Datos
- Responsabilidades y competencias de un Ingeniero de Datos
- Pantallazo de algunas de las aplicaciones típicas y del software más usado

Modulo 2: El ecosistema de datos

- Tipos de datos
- Fuentes de datos
- Trabajando con datos y fuentes diversos
- Introducción a RDBMS
- Introducción a NoSQL
- Data Warehouses, Data Marts, y Data Lakes
- Consideraciones generales para la elección de repositorios de datos
- ETL, ELT, y el Data Pipeline: ¿qué es?, ¿qué herramientas se usan?
- Integración de datos: importancia de estrategias robustas
- Algunas herramientas para Big Data: Hadoop, HDFS, Spark, Hive

Modulo 3: El ciclo de vida en Ingeniería de Datos

- Consideraciones de seguridad
- Recolección y almacenamiento de datos
- Introducción a ”Data Wrangling”
- Análisis de datos
- Visualización de datos
- Gobernanza
- Compliance

Modulo 4: Bases de datos relacionales

- Introducción al formato de datos y arquitectura
- Big Query: Setup y queries más usadas
- Ejemplos de aplicación en Ingeniería de Datos

Modulo 5: Bases de Datos NoSQL

- Introducción al formato de datos y arquitectura
- MongoDB: Setup y queries más usadas
- Ejemplos de aplicación en Ingeniería de Datos

Modulo 6: Shell scripting

- Comandos básicos de Linux y su uso en Ingeniería de Datos
- Data Wrangling: limpieza y consolidación de los datos
- Ejemplos de aplicación a 'data wrangling' y procesamiento de datos

Modulo 7: Big Data en Ingeniería de Datos

- Introducción a Apache Spark
- Introducción a Hadoop
- Ejemplos de aplicación

Modulo 8: Introducción a ETL

- ETL y ELT: Fuentes de datos, consolidación, y criterios importantes para la planificación
- Data Wrangling
- Diseñando data pipelines: consideraciones importantes
- Data pipelines con Shell, Kafka y Airflow
- Ejemplos de aplicación en Ingeniería de Datos

Modulo 9: Visualización de datos

- Introducción a Tableau y Looker
- Tipos de visualización adecuadas para cada tipo de dato, contexto, y audiencia
- ¿Cómo contar historias con datos?
- Ejemplos de aplicación con Looker y Tableau

Modulo 10: Introducción a Machine Learning

- Definiciones y conceptos básicos
- La relación entre Ingeniería de Datos y Machine Learning: ciclo de vida en Machine Learning – Aplicaciones típicas de Machine Learning
- Preparación y análisis de datos teniendo en cuenta los objetivos buscados
- Algunos algoritmos importantes y cuál usar en cada caso
- ¿Cómo explicar los resultados de Machine Learning a distintas audiencias?



Aprender Online se reserva el derecho de modificar este programa, o cualquiera de los componentes del curso, cuando lo considere necesario a fin de generar mejoras o actualizaciones.