

DATA SCIENCE 2: Ingeniería de Datos

Descripción de curso

En el curso de Ingeniería de Datos, aprenderás los fundamentos y las técnicas necesarias para diseñar, gestionar y procesar grandes volúmenes de datos de manera eficiente y efectiva.

Comenzarás con el diseño de bases de datos, donde aprenderás a modelar conceptualmente los datos y convertirlos en un modelo relacional utilizando diagramas ER (Entidad-Relación). También estudiarás la normalización de bases de datos para garantizar la integridad y eficiencia en el almacenamiento de los datos.

Continuarás con la extracción y transformación de datos. Aprenderás técnicas para extraer datos de fuentes heterogéneas y realizar tareas de limpieza y preprocesamiento para garantizar la calidad de los datos. También explorarás técnicas de transformación y agregación de datos para obtener información relevante.

Luego, te adentrarás en el almacenamiento y gestión de datos. Estudiarás las bases de datos relacionales, como SQL, y las bases de datos NoSQL, como MongoDB y Cassandra, que son especialmente adecuadas para manejar grandes volúmenes de datos no estructurados. También explorarás el almacenamiento en la nube y el procesamiento de Big Data para hacer frente a los desafíos de escalabilidad y rendimiento.

NEWBERT

CENTRO DE ESTUDIOS NEWBERT

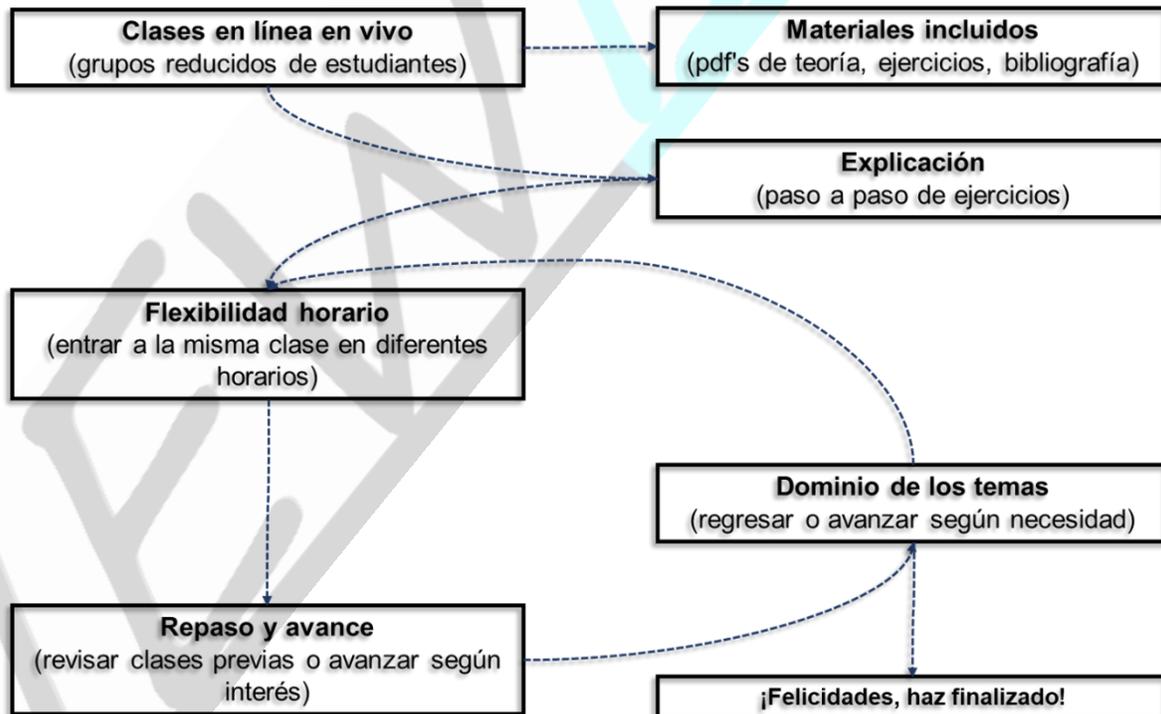
El procesamiento de datos a escala será otro tema importante en el curso. Aprenderás sobre el procesamiento en lotes utilizando tecnologías como Hadoop y MapReduce, que permiten procesar grandes volúmenes de datos de manera eficiente. También estudiarás el procesamiento en tiempo real utilizando herramientas como Apache Kafka y Apache Spark, que te permitirán analizar y responder a los datos en tiempo real.

A lo largo del curso, trabajarás en proyectos prácticos donde aplicarás los conceptos y las técnicas aprendidas utilizando herramientas populares en la industria como Python, SQL, Hadoop y Spark. Al finalizar el curso, estarás preparado para enfrentar los desafíos de la ingeniería de datos y serás capaz de diseñar, gestionar y procesar datos a gran escala de manera efectiva y eficiente.

Características del curso

1. Las clases se llevan a cabo de manera online en vivo en grupos reducidos de estudiantes.
2. Materiales incluidos (pdf's de la teoría vista en clase, ejercicios para practicar, bibliografía para consulta, entre otros).
3. Todos los ejercicios se explican paso a paso para su fácil entendimiento.
4. Se puede entrar a la misma clase en otro horario de la semana y las veces que se desee hasta dominar el tema.
5. Puedes avanzar clases si lo deseas o regresar a clases cuantas veces quieras para dominar los temas de interés, durante la vida del curso.

Metodología del curso



Contenido del curso

En este curso se cubren los siguientes temas:

Ingeniería de Datos

Diseño de Bases de Datos

- Modelado conceptual de datos
- Modelado relacional y diagramas ER
- Normalización de bases de datos

Extracción y Transformación de Datos

- Extracción de datos de fuentes heterogéneas
- Limpieza y preprocesamiento de datos
- Transformación y agregación de datos

Almacenamiento y Gestión de Datos

- Bases de datos relacionales (SQL)
- Bases de datos NoSQL (MongoDB, Cassandra)
- Almacenamiento en la nube y Big Data

Procesamiento de Datos a Escala

- Procesamiento en lotes (Hadoop, MapReduce)
- Procesamiento en tiempo real (Apache Kafka, Apache Spark)
- Streaming de datos y procesamiento por lotes