



# CURSO DATA ENGINEER

Plan de estudio



educación 





## Nuestro propósito

**Transformar positivamente la vida de las personas.**

Queremos que seas protagonista en la transformación que estamos viviendo. Por eso, nos comprometemos a capacitarte para que estés al día con las necesidades digitales actuales.

Te invitamos a trabajar en conjunto para que descubras tu mejor versión y la potencies. Anímate, toma las riendas de tu futuro.

Code your future!

## Contenido del curso

Conviértete en un Ingeniero de Datos creando pipelines con Spark, Python, PostgreSQL y Airflow, integrando clusterings de datos y arquitecturas Big Data.

## Proyecto Integrador

### Pipeline de datos profesional

En este trabajo aplicarás todo lo aprendido durante el curso para construir un pipeline completo y funcional. Realizarás:

- Limpieza, transformación y almacenamiento de datos.
- Ejecución en entornos reproducibles con Docker.
- Procesamiento masivo con Spark.
- Orquestación de flujos con Apache Airflow.
- Integración de bases SQL y NoSQL.
- Almacenamiento escalable en S3 / Minio.

Este proyecto, con casos de uso reales, formará parte de tu portfolio personal, y te servirá como experiencia profesional. Al finalizar, podrás compartir el diploma en LinkedIn para destacar tu perfil utilizando @educacionit desde tus redes.

## ¿Qué aprenderás?

- Procesos ETL/ELT.
- Modelado de datos: OLTP a OLAP.
- Streaming de eventos: Kafka.
- Python para procesamiento/limpieza de datos.
- Streaming vs. Batch.
- Arquitectura Lambda y Arquitectura Kappa.
- Docker para levantar entornos de práctica.
- Data Warehouse, Data Lake y Lakehouse.
- Almacenamiento en Object Storage (S3/Minio).
- Procesamiento de datos masivos.
- Bases de datos SQL y NoSQL,
- Orquestación de pipelines con Airflow.

# Plan de estudios

**1**

## Introducción y fundamentos de datos

- Rol del data engineer.
- Gobernanza de datos.
- Calidad de datos y estandarización.
- Preparación, integración y limpieza de dato.
- Sintaxis en Spark con Google Colab.
- Modelado dimensional.
- Estandarización de datos y buenas prácticas.
- Docker para levantar entornos de práctica.
- Introducción a orquestación de datos.

**2**

## Procesos ETL/ELT

- Implementación de una Base PostgreSQL.
- Análisis de extracción.
- Ingesta de datos
- APIs y Web Scraping.
- Extracción, transformación y carga en SQL.
- Ejecución de pipelines completos ETL y ELT.
- Diferencia entre bases OLTP y OLAP.
- Procesamiento de datos con Python (Pandas).
- Lab1: Comparativas de Bases relacionales.
- Lab2: Spark Dataframes.
- Lab3: Pandas Dataframes vs SparkSQL.

**3**

## Procesamiento de datos masivos

- Clustering de datos.
- Distribución de la carga de trabajo.
- Procesamiento distribuido.
- Fundamentos de Apache Spark.
- Librerías Spark Core, Spark SQL y Pyspark.
- Optimización del flujos de trabajos.
- Conexión entre Spark/procesamiento batch.
- Procesamiento masivo casi en tiempo real.
- Lab: procesar datos con Spark y Kafka.

## 4

### Arquitecturas de datos modernas

- Data Warehouse, Data Lake y Data Lakehouse.
- Streaming Vs Batch.
- Arquitectura Lambda y Arquitectura Kappa.
- Estructura y estrategias de integración.
- Almacenamiento en Nube S3/Minio.
- Object Storage (Minio) para persistencia.
- Contenedores y entornos reproducibles (Docker).

## 5

### Bases de datos SQL y NoSQL

- Conceptos ACID, BASE y Teorema CAP.
- Bases relacionales: PostgreSQL.
- Usos de SQL intermedio.
- Funciones ventana.
- Query plan.
- Evaluación de rendimiento de datos.
- Evaluación de calidad de datos
- Evaluación de consistencia de datos.

- Agregaciones.
- Condiciones
- Bases NoSQL: MongoDB, DynamoDB.
- Casos de uso e integración entre motores.

## 6

### Orquestación de pipelines (Airflow)

- Conceptos de workflows y DAGs.
- Instalación y configuración de Airflow.
- Ejecución de pipelines.
- Dependencias, scheduling y errores.
- Orquestación: Pandas, Spark, S3, PostgreSQL

## Modalidad del Curso

### Duración

8 semanas / 30 h

### Frecuencia semanal

2 encuentros de 2 h

### Modalidad

Online en vivo

### Grupos reducidos

Promedio 15 personas

## Nivel: Experto



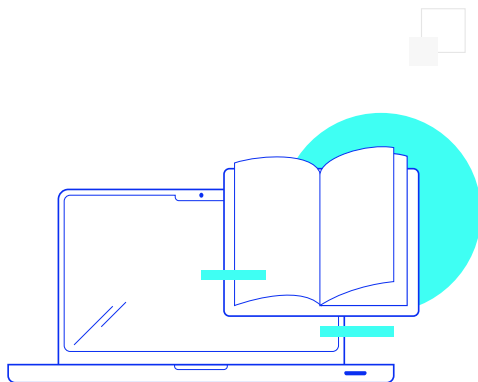
## Requisitos

Es recomendable tener una base sobre:

- **Introducción a Python**
- **Introducción a Base de Datos**
- **MongoDB Fundamentos**

## Dedicación fuera de clase

Además de las horas de clase, recomendamos que inviertas 4 h semanales extras para realizar los desafíos complementarios, estudiar el material de lectura y completar los exámenes del Alumni.



## ¿Cómo será tu experiencia?



### Aprender haciendo

Ejercita y pon en práctica lo estudiado.



### Trabajo en equipo

Une esfuerzos y potencia los resultados.



### Clases grabadas

Consúltalas las veces que quieras.



### Profesores expertos

Aprende de gigantes de la industria.



### Asistente académico

Recibe soporte dentro y fuera de clase.



### Plataforma Alumni

Encuentra recursos, materiales y clases.

## ¿Por qué Educación IT?



### IT Créditos

Gana puntos al aprobar los exámenes de los cursos. Luego, podrás canjearlos por nuevos cursos sin costo alguno. Los IT Créditos que acumules no vencen ni se devalúan.



### Garantía de aprendizaje

Si necesitas reforzar conceptos, recuperar clases o no estás satisfecho, ¡vuelve a tomar el curso sin ningún costo! Puede ser de forma total o parcial.



### Comunidad en Discord

Mantente en contacto con la comunidad de EducaciónIT a través de nuestro servidor de Discord. Podrás hablar con tus compañeros, profesores, asistentes académicos y soporte.

## Preguntas frecuentes

**Si me pierdo una o más clases, ¿puedo recuperarlas?**

Todas las clases quedan grabadas de por vida en tu plataforma Alumni. ¡Siempre podrás volver a verlas cada vez que lo necesites!

**¿Cómo voy a aprender?**

Te enfrentarás a situaciones de trabajo reales, en donde tendrás que aplicar lo aprendido de forma individual y en equipo. Por medio de la prueba y el error, irás superando desafíos y obteniendo nuevas habilidades que luego podrás aplicar en el ámbito laboral.

**¿Cómo son las clases online en vivo?**

Las clases duran entre 2 y 3 horas de lunes a viernes (sábados 3 o 4 hs) y se desarrollan de forma online en vivo en aulas virtuales, donde vas a poder interactuar con el instructor y tus compañeros.

Manejamos cupos reducidos para que puedas tener un seguimiento más personalizado durante tu aprendizaje.



[www.educacionit.com](http://www.educacionit.com)



@educacionit

---