



Vitae

 **CURSO**

Databricks para Ingeniería de Datos

21 Horas | 12, 14, 18, 20, 21, 26 y 28 de Noviembre de 2024

Introducción

Este curso de Databricks para Ingeniería de Datos está diseñado para profesionales que desean obtener una introducción del manejo de grandes volúmenes de datos utilizando una plataforma avanzada como Databricks.

Iremos desde los conceptos fundamentales hasta la implementación de soluciones complejas que abordan problemas reales de ingeniería de datos. Con laboratorios prácticos en cada módulo, aprenderás a dominar Apache Spark, automatizar pipelines de datos y optimizar los recursos en la nube. Al final del curso, estarás preparado para implementar soluciones eficientes y escalables en el mundo real, independientemente del tamaño de los datos que manejes.

Se grabarán las sesiones para poder consultarlas de forma ilimitada una vez finalizado el curso.

Objetivos :

- ✔ Comprender los conceptos básicos de la ingeniería de datos modernas basadas en arquitecturas Medallón.
- ✔ Aprender a utilizar la plataforma Databricks para la gestión y transformación de datos.
- ✔ Conocer la arquitectura Lakehouse y su aplicación en Databricks.
- ✔ Realizar transformaciones de datos utilizando PySpark.
- ✔ Implementar y gestionar tablas Delta.

Temario

✓ Introducción a la Ingeniería de Datos

- ¿Qué es la Ingeniería de datos?
- Conceptos clave: ETL, ELT, pipelines de datos, big data
- Introducción a Databricks y Apache Spark
- Arquitecturas de datos modernas
- Laboratorio: Configuración inicial de un entorno Databricks

✓ Fundamentos de Apache Spark

- Modelo de procesamiento distribuido de Spark
- Componentes clave: RDD, DataFrames, SparkSQL
- Transformaciones y acciones en Spark
- Operaciones comunes de manipulación de datos
- Laboratorio: Creación y manipulación de DataFrames en Spark

✓ Procesamiento de datos con Databricks

- Pipelines de datos: Conceptos y diseño
- Optimización de pipelines con Spark
- Gestión y Optimización de particiones
- Laboratorio: Construcción de un pipeline ETL con Spark

▣ Almacenamiento de datos y Delta Lake

- Introducción a Delta Lake
- Ventajas de Delta Lake sobre otros formatos
- Versionado de datos y transacciones ACID
- Ingestas de Datos
- Laboratorio: Implementación de Delta Lake para almacenamiento de datos.

Información del curso



Duración

21 horas lectivas



Modalidad

Aula Virtual con clases en directo y acceso a las sesiones grabadas para su consulta.



Fechas

12, 14, 18, 20, 21, 26 y 28 de Noviembre de 2024



Horarios

De 16:00 a 19:00 h



Dónde

Aula Virtual de Vitae



Formador

Antonio Soto

Ingeniero en Informática de Sistemas en la UDC

Profesional de más de 25 años de experiencia como Consultor experto y certificado, impartiendo cursos en la Universidad Microsoft y a empresas y organismos como LA NASA, NASDAQ, PEMEX, SERGAS, PSA - CITROEN, CEAGA, INDITEX, FERROPLAST, BANCO POPULAR, CAIXA GALICIA, XUNTA DE GALICIA, ABERTIS, MINISTERIO DE DEFENSA, BBC, SOGECABLE, GAS NATURAL, CAJA ESPAÑA, COMUNIDAD DE MADRID, COREMAIN, ALTIA CONSULTORES, BALIDEA, Universidad Politécnica de Madrid, UNIVERSIDAD DE VIGO, BBVA, SOLUTIONS30, etc...

Director Verne Technology Group

Consultor Asociado de Vitae Consultores

Condiciones económicas



Tarifa Por Asistente
260€ (Cuota bonificable por la FUNDAE)



Tarifa por asistente a partir de dos personas de la misma empresa u organización
230€ (Cuota bonificable por la FUNDAE)



Forma de Pago:
Por transferencia al finalizar el curso a la recepción de la factura
Se añadirá el 21 % de IVA



Inscripción:
vitae@vitaedigital.com
Tlf : 986 47 21 01
637 82 02 57