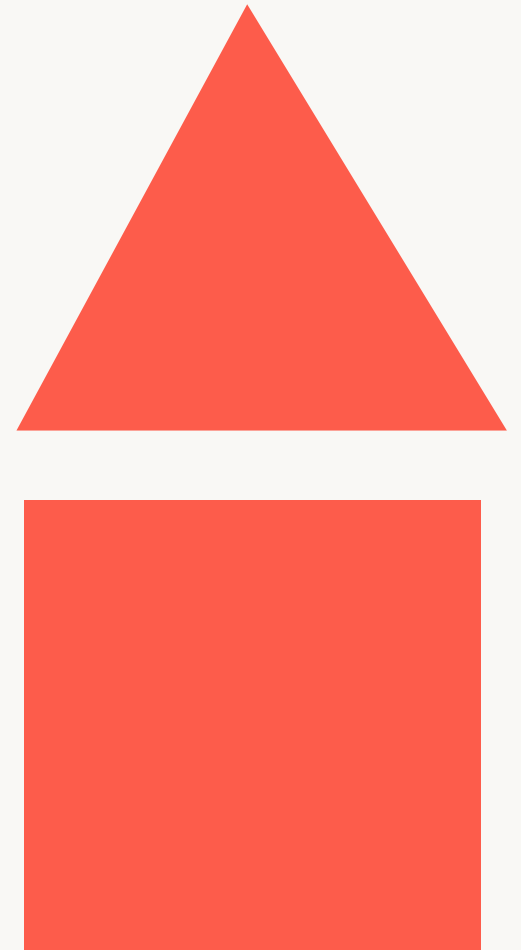
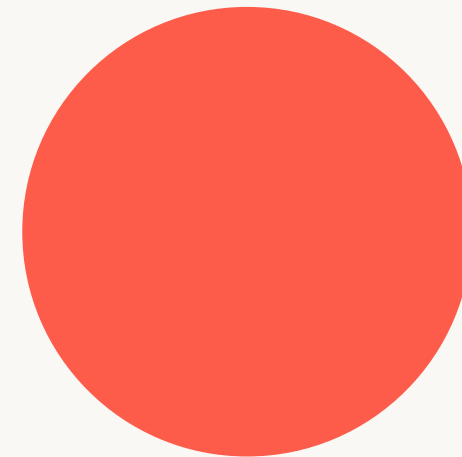


Curso: databricks

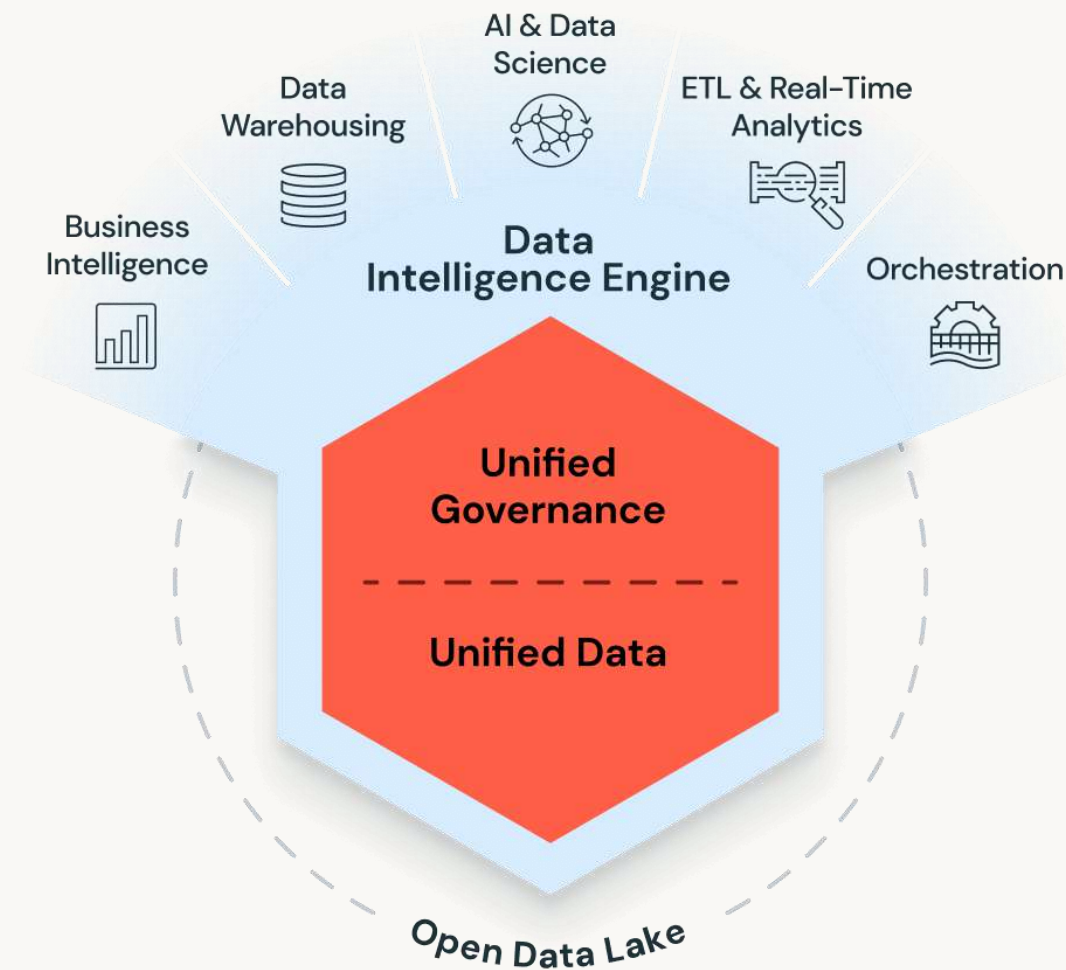


Domina la administración y operación de Databricks para potenciar el procesamiento de datos en la nube, mejorar el rendimiento de tus flujos de trabajo y aprovechar el poder de la inteligencia artificial y machine learning. Aprende a gestionar espacios de trabajo, seguridad, almacenamiento y automatización con las mejores prácticas del sector.

100% online - en vivo



¿Qué es Databricks?



Databricks es **una plataforma unificada** basada en la nube que facilita la **ingeniería de datos, el análisis avanzado y la inteligencia artificial**.

Construida sobre **Apache Spark**, permite a las organizaciones procesar y analizar **grandes volúmenes de datos** de forma eficiente.

Su entorno **colaborativo** favorece la cooperación entre analistas, ingenieros de datos y científicos de datos, lo que posibilita **desarrollar, escalar y gestionar pipelines de datos, modelos de machine learning y análisis en tiempo real** dentro de una misma plataforma.

Entre sus principales características se destacan la **integración nativa** con **Azure, AWS y Google Cloud**; la optimización del almacenamiento y la calidad de los datos mediante **Delta Lake**; y la automatización de procesos con **MLflow**, que gestiona el ciclo de vida de los modelos de inteligencia artificial.

Curso Databricks

- ✓ En nuestro curso conocerás la Administración y cómo poner en Operación los componentes de Databricks.
- ✓ Los participantes serán capaces de gestionar eficientemente entornos en la nube para el procesamiento de grandes volúmenes de datos
- ✓ Aprende a desarrollar modelos de Ciencia de Datos e Inteligencia Artificial para anticipar tendencias y optimizar procesos.
- ✓ Profundiza el aprendizaje con nuestros casos de uso, donde analizaremos las mejores prácticas para obtener los mejores resultados.

Dirigido a:

- Analista de BI
- Ingenieros de Datos
- Científicos de Datos
- DBAs

Metodología:

- ✓ 100% ONLINE EN VIVO – Guiado por Instructor



databricks



Módulo 1: Administración y Operación

10 horas de clase

1

Introducción a Databricks

- Conceptos claves y arquitectura de Databricks.

2

Administración de espacios de trabajo

- Configuración y gestión de clústeres
- Control de acceso y seguridad
- Roles y permisos de usuario
- Automatización y Monitoreo: Jobs y workflows

3

Integración y Despliegue continuo

- Construcción y automatización de workflows para pipelines de ingeniería y ciencia de Datos.

Módulo 2: Ingeniería de Datos

10 horas de clase

1

Delta Lake, Unity Catalog y Delta life tables

- Implementación de arquitectura en medallón para procesos en Batch y streaming

2

Transformaciones

- Transformaciones PySpark & SparkSQL
- Optimizaciones y buenas prácticas
- Construcción de Capa Semántica
- Optimizaciones y buenas prácticas

3

Construcción de pipelines

- Pipelines con Workflows

Módulo 3: Ciencia de Datos

10 horas de clase

1

Machine Learning

- Tratamiento de datos y EDA
- Machine Learning con MLlib
- Visualizaciones

2

Inteligencia Artificial con Copilot

- Gestiona el ciclo de vida de los modelos analíticos

Módulo 4: Casos

10 horas de clase

1

Casos prácticos

Introducción a los Casos de Uso en la Industria:

- Implementación técnica en Databricks para cada caso
- Prácticas y desarrollo de proyectos reales
- Conclusiones y mejores prácticas

Incluye

Certificado digital

Otorgado por entidad
autorizada de Databricks

Material

Código fuente y documentación
complementaria

Importante:

- El curso se realizará en un ambiente de Databricks del Cliente
- En el caso de requerir el Curso de Databricks de manera presencial, El Cliente deberá cubrir los gastos de traslado, tickets aéreos, viáticos, y todos los gastos logísticos del Curso