



OpenWebinars certifica que  
**Salvador Vázquez Roldán**

Ha superado con éxito  
**Dominando la Arquitectura Big  
Data con Databricks**

Duración del curso

**12 horas**

Fecha de expedición

**25 junio 2025**

A handwritten signature in black ink, appearing to read 'Manuel Agudo', positioned over a large, light grey watermark of the OpenWebinars logo.

CEO de OpenWebinars

**Manuel Agudo**

## Contenido

# Dominando la Arquitectura Big Data con Databricks

## 1. INTRODUCCIÓN

Presentación

## 2. FUNDAMENTOS Y RECOLECCIÓN DE DATOS

Introducción a la Arquitectura Big Data y Databricks

Introducción y arquitectura Spark

Recolección de datos en tiempo real

Práctica: Recolección de datos en tiempo real

Automatización de la carga y ETL con Databricks Auto Loader

Práctica: Automatización de la carga y ETL con Databricks Auto Loader

Test de Autoevaluación

## 3. PROCESAMIENTO DE DATOS CON SPARK

Procesamiento de datos con Spark en Databricks

Spark: Ejemplos de transformaciones y acciones con RDDs

Práctica Spark: Ejemplos de transformaciones y acciones con RDDs

SparkSQL: Transformaciones y acciones con DataFrames y Datasets

Práctica SparkSQL: Transformaciones y acciones con DataFrames y Datasets

Test de Autoevaluación

## 4. ALMACENAMIENTO DE DATOS

Introducción a Databricks Delta Lake

Organización y gestión de datos con Delta Lake

Práctica: Organización y gestión de datos con Delta Lake

Mejores prácticas de almacenamiento en Delta Lake



Práctica: Mejores prácticas de almacenamiento en Delta Lake

Test de Autoevaluación

## 5. PROCESAMIENTO ANALÍTICO CON SPARK MLlib

Spark MLlib: Librería de Machine Learning de Spark

Práctica Spark MLlib: Librería de Machine Learning de Spark

Discretizando variables en modelos analíticos con MLlib

Práctica: Discretizando variables en modelos analíticos con MLlib

Spark MLlib: Algoritmos basados en árboles

Práctica Spark MLlib: Algoritmos basados en árboles

Spark MLlib: Hiperparámetros y optimización

Práctica Spark MLlib: Hiperparámetros y optimización

Test de Autoevaluación

## 6. VISUALIZACIÓN Y CONSULTAS

Introducción a la visualización de datos en Databricks

Práctica: Introducción a la visualización de datos en Databricks

Optimización de consultas con Databricks SQL

Práctica: Optimización de consultas con Databricks SQL

Creación de dashboards interactivos en Databricks

Test de Autoevaluación

## 7. MEJORES PRÁCTICAS

Implementación y pruebas

Configuración de políticas de seguridad en Databricks

Estrategias para la optimización de consultas y procesos

Estrategias para la optimización de almacenamiento

Test de Autoevaluación

## 8. PROYECTO PRÁCTICO

Dashboard analítico con Databricks

## 9. CONCLUSIONES



Conclusiones y próximos pasos

