



Gracias a



openwebinars.net/cert/OI4T



OpenWebinars certifica que

ADRIÁN GAYO ANDRÉS

Ha superado con éxito

**Análisis de Big Data con Apache
Spark**

Duración del curso

8 horas

Fecha de expedición

06 mayo 2025

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

CEO de OpenWebinars

Manuel Agudo

Análisis de Big Data con Apache Spark

1. INTRODUCCIÓN A APACHE SPARK

Presentación

¿Qué es Apache Spark? Visión general y casos de uso

Configurando el entorno de desarrollo de Spark

RDD y DataFrames: Fundamentos y diferencias

Transformaciones y acciones básicas en Spark

Test de Autoevaluación

2. PROCESAMIENTO DE DATOS CON SPARK

Operaciones de conjuntos y su utilidad en análisis de datos

Trabajo con pares clave-valor para procesamiento avanzado

Técnicas de persistencia y particionado de datos

Depuración y optimización de trabajos en Spark

Test de Autoevaluación

3. MACHINE LEARNING CON SPARK

Introducción a MLlib y herramientas de Machine Learning

Implementación de algoritmos de aprendizaje supervisado

Exploración de algoritmos de aprendizaje no supervisado

Aplicaciones de sistemas de recomendación y análisis de grafos

Test de Autoevaluación

4. GESTIÓN DE RECURSOS Y OPTIMIZACIÓN

Técnicas avanzadas de optimización de consultas

Gestión eficiente de recursos y estrategias de particionado avanzado

Monitoreo y debugging de aplicaciones Spark en producción

Test de Autoevaluación

5. SPARK STREAMING Y ANÁLISIS EN TIEMPO REAL

Fundamentos de Spark Streaming

DStreams: Operaciones de E/S

Integración con sistemas de mensajería como Kafka

Implementación de análisis de flujos de datos en tiempo real

Test de Autoevaluación

6. PROYECTO PRÁCTICO

Análisis de sentimiento en tiempo real

7. CONCLUSIONES

Conclusiones y próximos pasos