



OpenWebinars certifica que

**Beatriz Galán Arribas**

Ha superado con éxito

**Conociendo PySpark**

Duración del curso

**7 horas**

Fecha de expedición

**15 abril 2024**

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**

## Contenido

# Conociendo PySpark

### 1. INTRODUCCIÓN

Presentación

### 2. INTRODUCCIÓN AL BIG DATA

Fundamentos del Big Data y ejemplos de su aplicación

Tipos y fuentes de datos

Tecnologías y conceptos complementarios

Test de autoevaluación

### 3. INTRODUCCIÓN A SPARK COMO FRAMEWORK DE BIG DATA

Ecosistema del Big Data y las tecnologías principales

Apache Hadoop

Apache Spark

Test de autoevaluación

### 4. CONOCIENDO PYSPARK

Repaso de Python y del entorno de trabajo

PySpark, conceptos, documentación e instalación

Conociendo PySpark

Test de autoevaluación

### 5. PYSPARK A FONDO

Funciones en PySpark

Analítica de datos con PySpark

Introducción a Apache Koala

Test de autoevaluación

### 6. MACHINE LEARNING CON PYSPARK



Fundamentos del Machine Learning

Tipos de aprendizaje y aplicación del ML

Ejemplo y laboratorio práctico

Test de autoevaluación

## **7. SPARK EN LA NUBE**

Conociendo el servicio de Databricks

Creación de clústeres y ejecución de archivos

Carga e ingesta de datos

Test de autoevaluación

## **8. CONCLUSIONES**

Conclusiones y próximos pasos

