



OpenWebinars certifica que

**Sofia Gonzalez**

Ha superado con éxito

**Curso Apache Spark: Procesando  
Big Data**

Duración del curso

**10 horas**

Fecha de expedición

**28 junio 2022**

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

# Curso Apache Spark: Procesando Big Data

## 1. INTRODUCCIÓN Y PREPARACIÓN DEL ENTORNO

Presentación del curso y profesor

Fundamentos de Apache Spark

Descarga e instalación

Jupyter notebook

## 2. SPARK SQL

RDDs

Transformaciones

Acciones

Ejercicios RDD: Ejercicio I

Ejercicios RDD: Ejercicio II

Ejercicios RDD: Ejercicio III

Persistencia y acumuladores

DataFrame y DataSet (parte I)

DataFrame y DataSet (parte II)

DataFrame y DataSet (parte III)

DataFrame y DataSet (parte IV)

Ejercicios con DataFrame y DataSet: Ejercicio I

Ejercicios con DataFrame y DataSet: Ejercicio II

Ejercicios con DataFrame y DataSet: Ejercicio III

Vistas

Ejercicio con Vistas

### 3. SPARK STREAMING

Structured Streaming (parte I)

Structured Streaming (parte II)

Structured Streaming (parte III)

Ventanas y watermark

Ejercicios con Streaming: Ejercicio I

Ejercicios con Streaming: Ejercicio II

Operación join y Monitorización

Spark Streaming DStream

Operaciones con ventana

Trabajando con Dataframe

### 4. PROYECTO CON TWITTER

Preparación del IDE

Ingestando tweets (parte I)

Ingestando tweets (parte II)

Ejercicios con Twitter: Ejercicio I

Ejercicios con Twitter: Ejercicio II

