



# CERTIFICACIÓN PYSPARK FUNDAMENTALS

**Let's code the knowledge of tomorrow**

## PYSPARK FUNDAMENTALS

# DOMINA LOS DATOS CON CONFIANZA

¿Quieres convertirte en un experto en procesamiento de datos a gran escala? En este curso de la mano de DaCodes Academy, aprenderás los fundamentos de PySpark, uno de los **frameworks** más utilizados para el **procesamiento de datos en clústeres**. Desde la **manipulación de datos** y el uso de **DataFrames** hasta la **optimización de tareas** y el **despliegue de pipelines** de datos.

[Inscribirme ahora](#)



Trusted Partners



Let's learn together | Inscríbete [aquí](#)

DEL CONOCIMIENTO A LA ACCIÓN

# CONOCE NUESTRO PROGRAMA DE CONTENIDOS

## Módulo 1: Introducción a PySpark

Desde la arquitectura de Spark hasta tu primer "Hello World" en PySpark:

- Introducción a Spark
- Arquitectura de Spark
- SparkSQL+Python: Pyspark
- Configuración de ambiente de desarrollo en docker y docker-compose
- Hello world: Lectura y escritura en pyspark

## Módulo 2: Inspección, proyección y filtrado de datos

Aprende a inspeccionar, proyectar, y filtrar datos eficazmente:

- Métodos: show(),count() y printSchema()
- Método select()
- Métodos: filter(),like(),between(),when() y otherwise()
- Ejemplo práctico
- Ejemplo práctico usando expr()
- Ejemplo práctico usando vistas

## Módulo 3: Modificación de columnas y agregación de datos

Domina las técnicas de modificación de columnas y funciones de agregación:

- Método withColumn()
- Método withColumnRenamed()
- Método drop()
- Métodos: agg(),groupBy() y orderBy()
- Ejemplo práctico
- Ejemplo práctico de modificación de columnas usando expr()
- Ejemplo práctico usando vistas

## Módulo 4: Cruce y unión de datos

Técnicas avanzadas para la unión y cruce de datasets:

- Cruce tipo: Left,Rigth,Outer - Join
- Cruce tipo: Anti Join
- Cruce tipo: Natural Join
- Cruce tipo : Cross Join
- Métodos: union() y unionByName()

## Módulo 4 : Transformaciones especiales

Explora transformaciones complejas incluyendo tipos de datos, strings, y fechas:

- Casteo de tipo de datos en pyspark
- Transformaciones en strings y expresiones regulares
- Transformaciones en tipo date y timestamp
- Transformaciones con tipo null

## Módulo 5 : Proyecto y examen final

- Demostración práctica de las habilidades en PySpark, desde la concepción hasta la implementación.
- Cada proyecto será evaluado individualmente, ofreciendo retroalimentación constructiva para asegurar que cada participante pueda identificar áreas de fortaleza y oportunidades de mejora.



## ¿QUÉ NECESITO PARA CURSAR PYSPARK?

- Conocimientos básicos de Python
- Familiaridad básica con SQL
- Entendimiento básico de Docker
- Buena conexión a internet.
- Ganas de superarse.

Trusted Partners



# CONOCE AL INSTRUCTOR



## DATA ENGINEER/DATA ARCHITECT

Maestro en Matemáticas Aplicadas & Computación por la FES Acatlán y trabajo como Data Engineer/Data Architect con más de 4 años de experiencia en la creación y gestión de data pipelines para procesos ETL, modelación, orquestación, linaje y aseguramiento de la calidad de los datos. Me especializo en arquitecturas de datos modernas como Data Lakes, Data Warehouses y Data Lakehouses, optimizando la disponibilidad y calidad de los datos para impulsar soluciones avanzadas en Analytics, Ciencia de Datos y AI.

## INSIGNIA DACODES ACADEMY

¡Un logro para tu carrera!

Al completar el curso satisfactoriamente, recibirás una insignia de certificación que valida tus habilidades en PySpark, la cual podrás añadir a tu perfil profesional y redes sociales.



## REQUISITOS PARA ACREDITAR LA CERTIFICACIÓN

- Participación activa en todas las sesiones teóricas y prácticas.
- Asistencia mínima del 80% de las sesiones en vivo.
- Completar y presentar el proyecto final, demostrando competencia en los conceptos y técnicas enseñados durante el curso.
- Superar un examen final que cubra todos los temas del curso y entregarlo en la fecha indicada.

ACELERA TU CARRERA

# BENEFICIOS DE CERTIFICARTE CON NOSOTROS



## Dominio de PySpark

Adquiere una comprensión profunda de PySpark.



## Experiencia Práctica

A través de proyectos reales y ejercicios aplicados.



## Crecimiento laboral

Desbloquea nuevas oportunidades laborales.



## Red de Profesionales

Conecta con otros profesionales del sector.

**¡PREPÁRATE PARA IMPULSAR TU CARRERA EN DATA ENGINEERING Y CONVIÉRTETE EN UN EXPERTO EN EL PROCESAMIENTO DISTRIBUIDO DE DATOS!**

[Inscribirme ahora](#)

Trusted Partners



Let's learn together | Inscríbete [aquí](#)