DATAPAO

INSTRUCTOR LED TRAINING

GEOSPATIAL ANALYTICS WITH DATABRICKS

LEVERAGE THE POWER OF GEOSPATIAL ANALYTICS FOR INFORMED DECISION-MAKING.

This training equips data professionals, interested in working with geospatial data, with the understanding of the most critical concepts, the latest toolkit and the hands-on skills needed to effectively harness the potential of geospatial analytics to drive informed decision-making across diverse industries and domains.

REGISTER NOW >

WHO IS IT FOR?

This course is intended for data professionals (Data Engineers, Data Scientists or Data Analysts) who are interested in working with geospatial data.

LEARNING OBJECTIVES

This course will provide data professionals with the knowledge and practical skills necessary for effectively leveraging the Databricks Lakehouse Platform to analyze geospatial data. Upon completing this course, learners will be able to:

Understand the most **critical concepts** of **geospatial data analysis**

Identify geospatial use cases such as optimal locations for business expansion, analysis of environmental changes, improvement of urban planning strategies

Manipulate and analyze geospatial data using languages and libraries such as Python, GeoPandas, Databricks Mosaic, and SQL on Databricks

Develop skills in **spatial data mining**, spatial **clustering**, and **scalable location intelligence**

Leverage the Databricks Lakehouse platform to **ingest, transform and visualize** large amounts of **geospatial data** efficiently

Effectively **communicate spatial patterns** and trends through **data visualization** techniques using Databricks, Matplotlib, and Kepler

COURSE OUTLINE

The course consists of three main modules covering the following topics:

MODULE 1

FUNDAMENTALS OF GEOSPATIAL DATA AND ANALYSIS

- Geospatial data types and file formats
- Geometries and transformations
- Reading, transforming and writing data with GeoPandas and Rasterio
- Perform analytics using GeoPandas

MODULE 2

SAMPLE USE CASE IMPLEMENTATION

- Sample use case: identifying green areas in cities
- Reading and visualizing satellite data
- Analyzing NDVI map data
- Performing spatial clustering

MODULE 3

SCALING OF GEOSPATIAL WORKLOADS

- Spatial indexing using H3
- Distributed computing using Apache Spark
- Introduction to the Mosaic library
- Visualizing data with Databricks
- Applying the Lakehouse principle with geospatial data

PREREQUISITES

A certain technical background and familiarity with related concepts is required for course participation, including:

- Basic programming experience with Python
- Basic knowledge of the Pandas library
- Familiarity with cloud computing concepts
- Familiarity with basic SQL concepts

TRAINING FORMAT

This is a live, instructor-led training using hands-on exercises. The course is designed to take 2 full working days, however, it can be delivered in smaller chunks, e.g. in 4 half-day sessions. The course will be delivered with a mix of:

DEMOS

Prepared courseware that learners can follow along on their own environment as the instructor demonstrates its content

LABS

Prepared exercises, where learners need to work on their own to practice the learned skills

PRESENTATIONS

Slide deck with content to illustrate the learning material

INSTRUCTORS

- The instructor-led course is delivered by DATAPAO's expert technical instructors.
- All our instructors are certified data engineers/data scientists with extensive hands-on experience in data projects as well as teaching:

200+

TRAININGS DELIVERED EVERY YEAR

10+

CERTIFICATION CATEGORIES OWNED (DATABRICKS AND MICROSOFT AZURE)

10+

COURSE TYPES TAUGHT

REGISTER NOW >