
Python and Spark for Big Data (PySpark) Eğitimi

Eğitim Hakkında

Python is a high-level programming language famous for its clear syntax and code readability. Spark is a data processing engine used in querying, analyzing, and transforming big data. PySpark allows users to interface Spark with Python.

In this instructor-led, live training, participants will learn how to use Python and Spark together to analyze big data as they work on hands-on exercises.

Neler Öğreneceksiniz

- Learn how to use Spark with Python to analyze Big Data.
- Work on exercises that mimic real world cases.
- Use different tools and techniques for big data analysis using PySpark.

Eğitim İçeriği

Learning the Basics of Python Programming

- Getting Started with Python
- Using the Jupyter Notebook
- Using Variables and Simple Data Types
- Working with Lists
- Using if Statements
- Using User Inputs
- Working with while Loops
- Implementing Functions
- Working with Classes
- Working with Files and Exceptions
- Working with Projects, Data, and APIs

Learning the Basics of Spark DataFrame

- Getting Started with Spark DataFrames
- Implementing Basic Operations with Spark
- Using Groupby and Aggregate Operations
- Working with Timestamps and Dates

Working on a Spark DataFrame Project Exercise

Understanding Machine Learning with MLlib

Working with MLlib, Spark, and Python for Machine Learning

Understanding Regressions

- Learning Linear Regression Theory
- Implementing a Regression Evaluation Code
- Working on a Sample Linear Regression Exercise
- Learning Logistic Regression Theory
- Implementing a Logistic Regression Code
- Working on a Sample Logistic Regression Exercise

Understanding Random Forests and Decision Trees

- Learning Tree Methods Theory
- Implementing Decision Trees and Random Forest Codes
- Working on a Sample Random Forest Classification Exercise

Working with K-means Clustering

- Understanding K-means Clustering Theory
- Implementing a K-means Clustering Code
- Working on a Sample Clustering Exercise

Working with Recommender Systems

Implementing Natural Language Processing

- Understanding Natural Language Processing (NLP)
- Overview of NLP Tools
- Working on a Sample NLP Exercise

Streaming with Spark on Python

Overview Streaming with Spark



Sample Spark Streaming Exercise