

Big Data with Python

Duration: 3 Days

Description

Big Data and Python Applications course is a training program that focuses on big data technologies and Python programming language and aims to provide participants with theoretical knowledge and practical skills in big data analysis and processing.

This 3-day training aims to provide participants with the ability to understand big data technologies and use these technologies effectively with the Python programming language. Learning is generally provided through practical applications, project studies and sample scenarios, so that participants can develop the ability to use their theoretical knowledge in real-world applications. This training is very valuable for individuals who aim to pursue a career in fields such as data science, artificial intelligence, big data analysis, or professionals who want to improve their skills in these fields.

Outline

Overview of Big Data

- What is Big Data?
- Big Data Components and Characteristic Features
- Applications of Big Data – Real-Life Examples
- Why Big Data?

Roadmap for Professionals Working in Big Data

- Expected Skills for Professionals in Big Data Field
- Skills Required for Big Data Expertise

Big Data Technologies and Tools

- Technologies Used in Big Data Architecture
- Apache Hadoop Ecosystem
- Apache Spark Technologies
- Other Technologies

Tools and Installations for Big Data Application Development

Development Model for Big Data Applications in Distributed Architectures

Python Application Development Environment

- Anaconda
- Jupyter Notebook
- Google Colab

Big Data Application Development Processes with PySpark

- PySpark Installation
- Basic Data Frame Operations with PySpark
- SQL Operations with PySpark
- Big Data Visualization with PySpark

Machine Learning Applications in Big Data with PySpark

- Individual Activity Analysis (Theory, Model, Prediction)
 - Data Preprocessing
 - Machine Learning on Big Data with Logistic Regression Algorithm
- Customer Churn Analysis (Theory, Model, Prediction)
 - Data Preprocessing
 - Machine Learning with Gradient Boosting Machine (GBM) Algorithm
- Hotel Click Data Analysis (Theory, Model, Prediction)
 - Data Preprocessing
 - Machine Learning on Big Data with Random Forest Algorithm