



WhatsApp: +91-7530088009

Call: India +91-444-631-1234

USA +1-650-265-2492

Email: training@sparkdatabox.com

Web: <https://sparkdatabox.com>

Spark Databox is known to be a pre-eminent platform for software certificate training and career development. Through our mission, we are a pioneer or positive change, improve productivity, increase the workforce, and creating a career opportunity for everyone. We are India's #1 software training institute. Apart from receiving excellent live training you will also receive free self-paced video courses, training materials, placement support, mock interviews and many more.

Program Key Features

- ❖ 100% Practical training
- ❖ Experienced Trainers
- ❖ 100% Placement assistance
- ❖ Small batch size
- ❖ Customized training content
- ❖ Real-time project training
- ❖ Fully equipped cloud lab
- ❖ 100% Customer support
- ❖ 100% Money back guarantee

About Course:

In this [PySpark online course](#), you will discover how to utilize Spark from Python. Spark is a tool for managing parallel computation with massive datasets, and it integrates excellently with Python. PySpark is the Python unit that performs the rapture happens. Spark Databox online training course is intended to equip you with the expertise and experiences that are needed to become a thriving Spark Developer using Python. During the PySpark Training, you will gain an in-depth understanding of Apache Spark and the Spark Ecosystem, which covers Spark RDD, Spark SQL, Spark MLlib, and Spark Streaming. You will also obtain extensive knowledge of Python Programming language, HDFS, Sqoop, Flume, Spark GraphX, and Messaging System.

Contents

Section 1: Big Data Analytics introduction

- Big Data overview
- Characteristics of Apache Spark
- Users and Use Cases of Apache Spark
- Job Execution Flow and Spark Execution
- Complete Picture of Apache Spark
- Why Spark with Python
- Apache spark Architecture
- Big Data Analytics in industry

Section 2: Using Hadoop's Core: HDFS and MapReduce

- HDFS: What it is, and how it works
- MapReduce: What it is, and how it works
- How MapReduce distributes processing
- HDFS commands

Section 3: Spark Databox Cloud Lab

- How to access Spark Databox cloud lab?
- Step by Step instruction to access cloud Big data Lab.

Section 4: Data analytics lifecycle

- Data Discovery
- Data Preparation
- Data Model Planning
- Data Model Building
- Data Insights

Section 5: python 3.0 (Crash Course)

- Environment Setup
- Decision Making
- Loops and Number
- Strings
- Lists
- Tuples
- Dictionary
- Date and Time
- Regex
- Functions
- Modules
- Files I/O
- Exceptions
- Multi-Threading
- Set
- Lamda Function

Section 6: PySpark

- Introduction to SparkContext
- Environment Setup
- Spark RDD
- spark Caching
- Common Transformations and Actions
- Spark Functions
- Key-Value Pairs
- Aggregate Functions
- Working with Aggregate Functions
- Joins in Spark
- Spark DataFrame

Section 7: Advanced Spark Programming

- Spark Shared Variables
- Custom Accumulator
- Spark and Fault Tolerance
- Broadcast variables
- Numeric RDD Operations
- Per-Partition Operations

Section 8: Running Spark jobs on Cluster

- Spark Runtime Architecture
- Spark Driver
- Executors
- Cluster Managers
- Connecting Spark To Different File System and Perform ETL ,(Extraction Transformation and Loading)
- Connecting Spark To DataBases and Perform ETL (Extraction Transformation and Loading)
- Spark StorageLevel
- Spark Serializers
- Spark-Submit and Cluster Explanation
- Performance Tuning

Section 9: PySpark Streaming at Scale

- Introduction to Spark Streaming
- PySpark Streaming with Apache Kafka
- Real-world Practical use cases
- Operations On Streaming Dataframes and Datasets
- Window Operations

Section 10: Real-world project training

- PySpark project environment setup
- Real-world PySpark project
- Project demonstration
- Expert evaluation and feedback

Section 11: You made it!!

- Spark Databox PySpark certification

- Interview preparation
- Mock interviews
- Resume preparation
- Knowledge sharing with industry experts
- Counseling to guide you to a right path in PySpark development career

You made it!!

Post completion of PySpark Online Course, a proper orientation for placements is done. With this training from experienced trainers, as professionals, you will be equipped with different proficiencies. This is a chance to open up and widen your prospects.

- Spark Databox PySpark course certification
- Interview preparation
- Mock interviews
- Resume preparation
- Knowledge sharing with industry experts
- Counseling to guide you to a right path in PySpark development career



www.sparkdatabox.com

© 2009-2019 - Spark Databox. All rights reserved.