

Pyspark Training

COURSE CONTENT

GET IN TOUCH



Multisoft Systems
B - 125, Sector - 2, Noida



(+91) 9810-306-956



info@multisoftsystems.com



www.multisoftsystems.com

About Multisoft

Train yourself with the best and develop valuable in-demand skills with Multisoft Systems. A leading certification training provider, Multisoft collaborates with top technologies to bring world-class one-on-one and certification trainings. With the goal to empower professionals and business across the globe, we offer more than 1500 training courses, which are delivered by Multisoft's global subject matter experts. We offer tailored corporate training; project Based Training, comprehensive learning solution with lifetime e-learning access, after training support and globally recognized training certificates.

About Course

Multisoft Systems offers an intensive PySpark training course designed to equip participants with the essential skills required to excel in Big Data processing and analytics. This course is structured to provide a comprehensive understanding of Apache Spark, with a particular focus on Python API, PySpark.

Module 1: Spark

- ✓ Spark Basics
- ✓ What is Apache Spark?
- ✓ Spark Installation
- ✓ Spark Configuration
- ✓ Spark Context
- ✓ Using Spark Shell

Module 2: Resilient Distributed Datasets (RDDs) – Features, Partitions, Tuning Parallelism

- ✓ Functional Programming with Spark
- ✓ Working with RDDs

Module 3: RDD Operations - Transformations and Actions

- ✓ Types of RDDs
- ✓ Key-Value Pair RDDs – Transformations and Actions
- ✓ Overview
- ✓ A Spark Standalone Cluster
- ✓ The Spark Standalone Web UI
- ✓ Executors & Cluster Manager
- ✓ Spark on YARN Framework
- ✓ Writing Spark Applications
- ✓ Building and Running a Spark Application
- ✓ Spark Job Anatomy
- ✓ Caching and Persistence
- ✓ RDD Lineage
- ✓ Caching Overview
- ✓ Distributed Persistence
- ✓ Resilient Distributed Datasets (RDDs)

- ✓ Parallelized Collections
- ✓ External Datasets
- ✓ PySpark Built-in Functions
- ✓ PySpark Datasources

Module 4: Spark SQL

- ✓ Introducing SparkSQL
- ✓ Dataframes in Spark
- ✓ Different Ways of Creating Dataframes
- ✓ Datasets and its applicability in Pyspark
- ✓ Hands on examples of dataframe

Module 5: Spark Structured Streaming

Module 6: PySpark DataFrame