

Spark Syllabus

This course has basic pre-requisite of Hadoop. Candidates are expected to have basic understanding of Hadoop components.

Spark And Scala

Module 0 - Scala

- **What is Scala?**
 - Why Scala for Spark?
 - Scala in other frameworks
 - Introduction to Scala REPL
 - Basic Scala operations
 - Variable Types in Scala
 - Control Structures in Scala
 - Foreach loop
 - Functions
 - Procedures
 - Collections in Scala- Array, ArrayBuffer, Map, Tuples, Lists, and more.

Module 1 - Spark Core

- **Introduction**
 - Introduction to big data,
 - Challenges with big data
 - Batch Vs. Real Time big data analytics
 - Batch Analytics - Hadoop Ecosystem Overview

Website: www.dw-learnwell.com

Contact: +91 8411002339/+91 7709292162

Email: info@dw-learnwell.com

Classroom

| Corporate

| Online

Spark Syllabus

- Real-time Analytics
- **What is Spark?**
 - Spark Ecosystem
 - Modes of Spark
 - Spark installation demo
 - Overview of Spark on a cluster
 - Spark Standalone cluster
 - Spark Web UI
 - Some configurations.
- **Components of Spark Unified stack**
 - Spark Streaming
 - MLlib
 - Core
 - Spark SQL
- **RDD - The core concept of Spark**
 - RDDs,
 - Transformations in RDD,
 - Actions in RDD,
 - Loading data in RDD,
 - Saving data through RDD,
 - Key-Value Pair RDD,
 - MapReduce and Pair RDD Operations
- **Scala and Python shell**
 - Word count example
- **Shared Variables with examples**

Website: www.dw-learnwell.com

Contact: +91 8411002339/+91 7709292162

Email: info@dw-learnwell.com

Classroom

| Corporate

| Online

Spark Syllabus

- **Submitting jobs in cluster**
- **Hands on examples**

Module 2 - Spark SQL

- **Overview**
 - Hive and Spark SQL architecture
 - sqlContext in spark sql
- **Dataframes API**
 - Understanding concept of data frame
 - Loading data in dataframe
 - Operations on dataframes.
- **Interaction with Hive**
- **Reading various data formats**
- **Hands on Examples**

Module 3 - Spark Streaming

- **Overview of streaming**
 - Spark Streaming Architecture,
 - First Spark Streaming Program,
 - Transformations in Spark Streaming,
 - checkpointing,
 - Parallelism level
 - **Introduction to queuing systems. Eg. Kafka**
 - **Hands on examples**
-

Website: www.dw-learnwell.com

Contact: +91 8411002339/+91 7709292162

Email: info@dw-learnwell.com

Classroom

| Corporate

| Online

Spark Syllabus

Module 4 - Spark MLlib

➤ **Supervised Learning**

- Classification - logistic regression, decision trees, random forests, naive Bayes
- Regression - linear least squares, Lasso, ridge regression, decision trees

➤ **Unsupervised learning :**

- Clustering - K-means, Gaussian Mixture

➤ **Dimensionality reduction**

- PCA

➤ **Hands on examples**

Projects

Note: Hands-on sessions will be conducted for all the topics mentioned above.