



DATA SCIENCE SYLLABUS

Module 1

Introduction to Scala

Learning Objectives – In this module, you will understand basic concepts of Scala, motives towards learning a new language and get your set-up ready.

Topics

- Why Scala?
- What is Scala?
- Introducing Scala
- Installing Scala
- Journey – Java to Scala
- First Dive – Interactive Scala
- Writing Scala Scripts – Compiling Scala Programs
- Scala Basics
- Scala Basic Types
- Defining Functions
- IDE for Scala, Scala Community

Module 2

Scala Essentials

Learning Objectives – In this module, you will learn essentials of Scala that are needed to work on it.

Topics

- Immutability in Scala – Semicolons
- Method Declaration, Literals
- Lists
- Tuples
- Options
- Maps
- Reserved Words
- Operators
- Precedence Rules
- If statements
- Scala For Comprehensions
- While Loops
- Do-While Loops
- Conditional Operators
- Pattern Matching
- Enumerations

Module 3

Traits and OOPs in Scala

Learning Objectives – In this module, you will understand implementation of OOPs concepts in Scala and use Traits as Mixins

Topics

- Traits Intro – Traits as Mixins
- Stackable Traits
- Creating Traits Basic OOPS – Class and Object Basics

- Scala Constructors
- Nested Classes
- Visibility Rules

Module 4

Functional Programming in Scala

Learning Objectives – In this module, you will understand functional programming know how for Scala.

Topics

- What is Functional Programming?
- Functional Literals and Closures
- Recursion
- Tail Calls
- Functional Data Structures
- Implicit Function Parameters
- Call by Name
- Call by Value

Module 5

Introduction to Big Data and Spark

Learning Objectives – In this module, you will understand what is Big Data, it's associated challenges, various frameworks available and will get the first hand introduction to Spark

Topics

- Introduction to Big Data
- Challenges with Big Data
- Batch Vs. Real Time Big Data Analytics
- Batch Analytics – Hadoop Ecosystem Overview
- Real Time Analytics Options, Streaming Data – Storm
- In Memory Data – Spark
- What is Spark?
- Modes of Spark
- Spark Installation Demo
- Overview of Spark on a cluster
- Spark Standalone Cluster

Module 6

Spark Baby Steps

Learning Objectives – In this module, you will learn how to invoke Spark shell and use it for various common operations.

Topics

- Invoking Spark Shell
- Loading a File in Shell
- Performing Some Basic Operations on Files in Spark Shell
- Building a Spark Project with sbt, Building and Running Spark Project with sbt
- Caching Overview, Distributed Persistence
- Spark Streaming Overview
- Example: Streaming Word Count

Module 7

Playing with RDDs

Learning Objectives – In this module, you will learn one of the building blocks of Spark – RDDs and related manipulations for implementing business logics.

Topics

- RDDs
- Transformations in RDD
- Actions in RDD
- Loading Data in RDD
- Saving Data through RDD

- Scala and Hadoop Integration Hands on

Module 8

Shark – When Spark meets Hive (Spark SQL)

Learning Objectives – In this module, you will see various offspring's of Spark like Shark, SparkSQL and Mlib. This session is primarily interactive for discussing industrial use cases of Spark and latest developments happening in this area.

Topics

- Why Shark?
- Installing Shark
- Running Shark
- Loading of Data
- Hive Queries through Spark
- Testing Tips in Scala
- Performance Tuning Tips in Spark
- Shared Variables: Broadcast Variables
- Shared Variables: Accumulators

Module 9

Spark Streaming

Learning Objectives – In this module, you will learn about the major APIs that Spark offers. You will get an opportunity to work on Spark streaming which makes it easy to build scalable fault-tolerant streaming applications.

Topics

- Spark Streaming Architecture
- First Spark Streaming Program
- Transformations in Spark Streaming
- Fault tolerance in Spark Streaming
- Check pointing
- Parallelism level

Module 10

Spark Mlib

Learning Objectives – In this module, you will learn about the machine learning concepts in Spark

Topics

- Classification Algorithm
- Clustering Algorithm
- Sequence Mining Algorithm
- Collaborative filtering

Module 11

Spark GraphX

Learning Objectives – In this module, you will learn about Graph Analysis concepts in Spark

Topics

- Graph analysis with Spark
- GraphX for graphs
- Graph-parallel computation

Module 12

Project and Installation

Topics

- Installation of Spark and Scala
- Discussion of real time use cases using Spark
- Mini project implementation in Spark