# Apache Spark and Scala Certification Training

**IntelliPaat**

## About Intellipaat

Intellipaat is a fast-growing professional training provider that is offering training in over 150 most sought-after tools and technologies. We have a learner base of 600,000 in over 32 countries and growing. For job assistance and placement we have direct tie-ups with 80+ MNCs.

**Key Features of Intellipaat Training:**

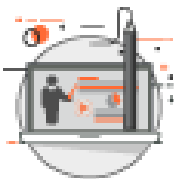| | | | |
|---|---|---|---|
| **Instructor Led Training**<br>24 Hrs of highly interactive instructor led training | **Self-Paced Training**<br>22 Hrs of Self-Paced sessions with Lifetime access | **Exercise and project work**<br>60 Hrs of real-time projects after every module | **Lifetime Access**<br>Lifetime access and free upgrade to latest version |
| **Support**<br>Lifetime 24*7 technical support and query resolution | **Get Certified**<br>Get global industry recognized certifications | **Job Assistance**<br>Job assistance through 80+ corporate tie-ups | **Flexi Scheduling**<br>Attend multiple batches for lifetime & stay updated. |

## About the Course

Intellipaat Apache Spark and Scala Certification Training Course offer you the hands-on knowledge to create Spark applications using Scala programming. It gives you a clear comparison between Spark and Hadoop. The course provides you techniques to increase application performance and enable high-speed processing using Spark RDDs as well as help in customization of Spark using Scala.

**Instructor Led**
**Duration – 24 Hrs**
**Weekend Batch –3 Hrs/Session**

**Self Paced**
**Duration – 22 Hrs**

# Apache Spark and Scala Certification Training

**IntelliPaat**

## Why take this Course?

- ❖ Apache Spark is an open source computing framework up to 100 times faster than MapReduce
- ❖ Spark is an alternative form of data processing unique in batch processing and streaming
- ❖ This is a comprehensive course for advanced implementation of Scala
- ❖ Prepare yourself for cloud era Hadoop Developer and Spark Professional Certification
- ❖ Get professional credibility to your resume so you get hired faster with high salary



## Course Curriculum

**Scala Course Content:**

| Module /Topic |
| --- |
| **Introduction to Scala**<br><br>   ❖ Introducing Scala and deployment of Scala for Big Data applications and Apache Spark analytics |
| **Pattern Matching**<br><br>   ❖ The importance of Scala<br>   ❖ The concept of REPL (Read Evaluate Print Loop)<br>   ❖ Deep dive into Scala pattern matching, type interface, higher-order function, currying, traits, application space and Scala for data analysis |
| **Executing the Scala code**<br><br>   ❖ Learning about the Scala Interpreter<br>   ❖ Static object timer in Scala |

❖ Testing String equality in Scala
❖ Implicit classes in Scala
❖ The concept of currying in Scala, various classes in Scala.

## Classes concept in Scala

❖ Learning about the Classes concept
❖ Understanding the constructor overloading
❖ The various abstract classes, the hierarchy types in Scala
❖ The concept of object equality, the val and var methods in Scala

## Case classes and pattern matching

❖ Understanding Sealed traits, wild, constructor, tuple, variable pattern, and constant pattern

## Concepts of traits with an example

❖ Understanding traits in Scala, the advantages of traits, linearization of traits
❖ The Java equivalent and avoiding of boilerplate code

## Scala Java Interoperability

❖ Implementation of traits in Scala and Java, handling of multiple traits extending

## Scala collections

❖ Introduction to Scala collections
❖ Classification of collections, the difference between Iterator, and Iterable in Scala, an example of list sequence in Scala

## Mutable collections vs. Immutable collections

❖ The two types of collections in Scala
❖ Mutable and Immutable collections
❖ Understanding lists and arrays in Scala, the list buffer and array buffer
❖ Queue in Scala, double-ended queue Deque, Stacks, Sets, Maps, Tuples in Scala

## Use Case bobsrockets package

❖ Introduction to Scala packages and imports

- ❖ The selective imports, the Scala test classes,
- ❖ Introduction to JUnit test class, JUnit interface via JUnit 3 suite for Scala test, packaging of Scala applications in Directory Structure
- ❖ An example of Spark Split and Spark Scala

## Spark Course Content

| Module /Topic |
|---|
| **Introduction to Spark** <br><br> ❖ Introduction to Spark <br> ❖ How Spark overcomes the drawbacks of working MapReduce <br> ❖ Understanding in-memory MapReduce, interactive operations on MapReduce, Spark stack <br> ❖ Fine vs. coarse-grained update, Spark stack <br> ❖ Spark Hadoop YARN, HDFS Revision, YARN Revision, the overview of Spark and how it is better Hadoop <br> ❖ Deploying Spark without Hadoop, Spark history server, Cloudera distribution |
| **Spark Basis** <br><br> ❖ Spark installation guide, Spark configuration <br> ❖ Memory management, executor memory vs. driver memory <br> ❖ Working with Spark Shell <br> ❖ The concept of Resilient Distributed Datasets (RDD) <br> ❖ Learning to do functional programming in Spark, the architecture of Spark |
| **Working with RDDs in Spark** <br><br> ❖ Spark RDD, creating RDDs, RDD partitioning <br> ❖ Operations & transformation in RDD <br> ❖ Deep dive into Spark RDDs <br> ❖ The RDD general operations <br> ❖ A read-only partitioned collection of records <br> ❖ Using the concept of RDD for faster and efficient data processing <br> ❖ RDD action for Collect, Count, Collects map, Save as textiles, pair RDD functions |

# Apache Spark and Scala Certification Training

**IntelliPaat**

## Aggregating Data with Pair RDDs

- ❖ Understanding the concept of Key-Value pair in RDDs
- ❖ Learning how Spark makes MapReduce operations faster
- ❖ Various operations of RDD
- ❖ MapReduce interactive operations, fine & coarse-grained update, Spark stack

## Writing and Deploying Spark Applications

- ❖ Comparing the Spark applications with Spark Shell
- ❖ Creating a Spark application using Scala or Java
- ❖ Deploying a Spark application, Scala built an application
- ❖ Creation of mutable list, set & set operations, list, tuple, concatenating a list
- ❖ Creating an application using SBT,
- ❖ Deploying application using Maven
- ❖ The web user interface of Spark application, a real-world example of Spark and configuring of Spark

## Parallel Processing

- ❖ Learning about Spark parallel processing
- ❖ Deploying on a cluster
- ❖ Introduction to Spark partitions, file-based partitioning of RDDs
- ❖ Understanding of HDFS and data locality
- ❖ Mastering the technique of parallel operations
- ❖ Comparing repartition & coalesce RDD actions

## Spark RDD Persistence

- ❖ The execution flow in Spark
- ❖ Understanding the RDD persistence overview
- ❖ Spark execution flow & Spark terminology
- ❖ Distribution of shared memory vs. RDD, RDD limitations
- ❖ Spark shell arguments, distributed persistence, RDD lineage
- ❖ Key/Value pair for sorting implicit conversion like Count By Key, Reduce By Key, Sort By Key, Aggregate By Key

## Spark Streaming and Mlib

- ❖ Spark Streaming Architecture
- ❖ Writing streaming program coding, processing of spark stream

- ❖ Processing Spark Discretized Stream (DStream)
- ❖ The context of Spark Streaming, streaming transformation
- ❖ Flume Spark streaming, request count and Dstream
- ❖ Multi-batch operation, sliding window operations, and advanced data sources
- ❖ Different Algorithms, the concept of the iterative algorithm in Spark, analyzing with Spark graph processing
- ❖ Introduction to K-Means and machine learning, various variables in Spark like shared variables, broadcast variables, learning about accumulators

## Improving Spark Performance

- ❖ Introduction to various variables in Spark like shared variables, broadcast variables
- ❖ Learning about accumulators
- ❖ The common performance issues and troubleshooting the performance problems

## Spark SQL and Data Frames

- ❖ Learning about Spark SQL, the context of SQL in Spark for providing a structured data processing
- ❖ JSON support in Spark SQL, working with XML data, parquet files
- ❖ Creating Hive Context, writing Data Frame to Hive, reading JDBC files
- ❖ Understanding the Data Frames in Spark
- ❖ Creating Data Frames, manual inferring of the schema, working with CSV files, reading JDBC tables, Data Frame to JDBC
- ❖ User-defined functions in Spark SQL shared variable and accumulators
- ❖ Learning to query and transform data in Data Frames
- ❖ How Data Frame provides the benefit of both Spark RDD and Spark SQL
- ❖ Deploying Hive on Spark as the execution engine

## Scheduling/ Partitioning

- ❖ Learning about the scheduling and partitioning in Spark, hash partition, range partition, scheduling within and around applications
- ❖ Static partitioning, dynamic sharing, fair scheduling, Map partition with an index
- ❖ The Zip, Group By Key, Spark master high availability
- ❖ Standby Masters with Zookeeper
- ❖ Single Node Recovery With Local File System
- ❖ High Order Functions

# Apache Spark and Scala Certification Training

## Project Work

### Project 1: Movie Recommendation

**Topics:** This is a project wherein you will gain hands-on experience in deploying Apache Spark for the movie recommendation. You will be introduced to the Spark Machine Learning Library, a guide to MLlib algorithms and coding which is a machine learning library. Understand how to deploy collaborative filtering, clustering, regression, and dimensionality reduction in MLlib. Upon completion of the project, you will gain experience in working with streaming data, sampling, testing and statistics.

### Project 2: Twitter API Integration for tweet Analysis

**Topics:** With this project, you will learn to integrate Twitter API for analyzing tweets. You will write codes on the server side using any of the scripting languages like PHP, Ruby or Python, for requesting the Twitter API and get the results in JSON format. You will then read the results and perform various operations like aggregation, filtering and parsing as per the need to come up with tweet analysis.

### Project 3: Data Exploration Using Spark SQL – Wikipedia data set

**Topics**: This project lets you work with Spark SQL. You will gain experience in working with Spark SQL for combining it with ETL applications, real-time analysis of data, performing batch analysis, deploying machine learning, creating visualizations and processing of graphs.

## Intellipaat Job Assistance Program

Intellipaat is offering comprehensive job assistance to all the learners who have successfully completed the training. A learner will be considered to have successfully completed the training if he/she finishes all the exercises, case studies, projects and gets a minimum of 60% marks in the Intellipaat qualifying exam.

Intellipaat has exclusive tie-ups with over 80 MNCs for placement. All the resumes of eligible candidates will be forwarded to the Intellipaat job assistance partners. Once there is a relevant opening in any of the companies, you will get a call directly for the job interview from that particular company.

# Apache Spark and Scala Certification Training

## Frequently Asked Questions:

**Q 1. What is the criterion for availing the Intellipaat job assistance program?**

Ans. All Intellipaat learners who have successfully completed the training post April 2017 are directly eligible for the Intellipaat job assistance program.

**Q 2. Which are the companies that I can get placed in?**

Ans. We have exclusive tie-ups with MNCs like **Ericsson, Cisco, Cognizant, Sony, Mu Sigma, Saint-Gobain, Standard Chartered, TCS, Genpact, Hexaware**, and more. So you have the opportunity to get placed in these top global companies.

**Q 3. Does Intellipaat help learners to crack the job interviews?**

Ans. Intellipaat has an exclusive section which includes the top interview questions asked in top MNCs for most of the technologies and tools for which we provide training. Other than that our support and technical team can also help you in this regard.

**Q 4. Do I need to have prior industry experience for getting an interview call?**

Ans. There is no need to have any prior industry experience for getting an interview call. In fact, the successful completion of the Intellipaat certification training is equivalent to six months of industry experience. This is definitely an added advantage when you are attending an interview.

**Q 5. What is the job location that I will get?**

Ans. Intellipaat will try to get you a job in your same location provided such a vacancy exists in that location.

**Q 6. Which is the domain that I will get placed in?**

Ans. Depending on the Intellipaat certification training you have successfully completed, you will be placed in the same domain.

**Q 7. Is there any fee for the Intellipaat placement assistance?**

Ans. Intellipaat does not charge any fees as part of the placement assistance program.

# Apache Spark and Scala Certification Training

**Q 8. If I don't get a job in the first attempt, can I get another chance?**

Ans. Definitely, yes. Your resume will be in our database and we will circulate it to our MNC partners until you get a job. So there is no upper limit to the number of job interviews you can attend.

**Q 9. Does Intellipaat guarantee a job through its job assistance program?**

Ans. Intellipaat does not guarantee any job through the job assistance program. However, we will definitely offer you full assistance by circulating your resume among our affiliate partners.

**Q 10. What is the salary that I will be getting once I get the job?**

Ans. Your salary will be directly commensurate with your abilities and the prevailing industry standards.

## What makes us who we are?



*"The Intellipaat Spark trainers were of the highest experience and knowledge. But what I liked was that the trainers went out their way in order to explain things and that too with real world examples which helped to learn Spark quickly. If I did not understand the first time they had the patience to explain it again making this learning experience one of a kind. The entire training is in line with clearing the Apache Spark Certification."*

*-Ibraheem Fayemi*



*"Spark training at Intellipaat ticked all the right boxes. The best thing I liked about the Apache Spark certification training was the opportunity to work on real world projects helping me get hands-on experience in one of the fastest big data processing engines. Thank you Intellipaat."*

*-Atyant Jain*

# Apache Spark and Scala Certification Training

*"I firmly believe Intellipaat is the perfect place to embark on a great professional career in the technology space. Their Apache Spark, Scala training course was praiseworthy."*

*-Suman*