APACHE SPARK AND SCALA CERTIFICATION TRAINING

About Intellipaat

Intellipaat is a global online professional training provider. We are offering some of the most updated, industry-designed certification training programs in the domains of Big Data, Data Science & AI, Business Intelligence, Cloud, Blockchain, Database, Programming, Testing, SAP and 150 more technologies.

We help professionals make the right career decisions, choose the trainers with over a decade of industry experience, provide extensive hands-on projects, rigorously evaluate learner progress and offer industry-recognized certifications. We also assist corporate clients to upskill their workforce and keep them in sync with the changing technology and digital landscape.

About The Course

Intellipaat is offering the comprehensive AWS certification training created by industry experts. The entire AWS training course is in line with the AWS Certified Solutions Architect exam. You will learn various aspects of AWS like Elastic Cloud Compute, Simple Storage Service, Virtual Private Cloud, Aurora database service, Load Balancing, Auto Scaling and more by working on hands-on projects and case studies. You will implement AWS best practices in this training.

Instructor Led Training
24 Hrs of highly interactive instructor led training

Self-Paced Training
22 Hrs of Self-Paced sessions with Lifetime access

Exercise and project work
60 Hrs of real-time projects after every module

Lifetime Access
Lifetime access and free upgrade to latest version

Support
Lifetime 24*7 technical support and query resolution

Get Certified
Get global industry recognized certifications

Job Assistance
Job assistance through 80+ corporate tie-ups

Flexi Scheduling
Attend multiple batches for lifetime & stay updated.

Why take this Course?

- Apache Spark is an open-source computing framework that is up to 100 times faster than MapReduce

- Spark is an alternative form of data processing, which is unique in batch processing and streaming

- This is a comprehensive course for advanced implementation of Scala

IntelliPaat

- It helps you prepare yourself for Cloudera Hadoop Developer and Spark Professional Certification
- Get professional credibility to your resume so that you get hired faster with high salary

---

Course Content

Scala

1. Introduction to Scala
2. Pattern Matching
3. Executing the Scala code
4. Classes concept in Scala
5. Case classes and Pattern Matching
6. Concept of Traits with example
7. Scala Java Interoperability
8. Scala Collections
9. Mutable Collection Vs. Immutable collection
10. Use case Bobsrockets Package

Spark

1. Introduction to Spark
2. Spark Basics
3. Working with RDDs in Spark
4. Aggregating Data with Pair RDDs
5. Writing and Deploying Spark Applications
6. Parallel Applications
7. Spark RDD Persistence
8. Spark MLlib
9. Integrating Apache Flume and Apache Kafka
10. Spark Streaming
11. Improving Spark Performance
12. Spark SQL and Data Frames
13. Scheduling/ Partitioning

Introduction to Scala

- ❖ Introducing Scala and deployment of Scala for Big Data applications and Apache Spark analytics

- ❖ Scala REPL and Lazy Values

- ❖ Control Structures in Scala

- ❖ Directed Acyclic Graph (DAG)

- ❖ First Spark Application Using SBT/Eclipse

- ❖ Spark Web UI and Spark in Hadoop Ecosystem.

Pattern Matching

- ❖ The importance of Scala

- ❖ The concept of REPL (Read Evaluate Print Loop)

- ❖ Deep dive into Scala pattern matching

- ❖ Type interface, higher-order function, currying, traits

- ❖ Application space and Scala for data analysis

Executing the Scala Code

- ❖ Learning about the Scala Interpreter

- ❖ Static object timer in Scala and testing string equality in Scala

- ❖ Implicit classes in Scala

- ❖ The concept of currying in Scala and various classes in Scala

Classes Concept in Scala

- ❖ Learning about the Classes concept

- ❖ Understanding the constructor overloading

- ❖ Various abstract classes and the hierarchy types in Scala

- ❖ The concept of object equality and 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 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 and handling of multiple traits extending

Scala Collections

❖ Introduction to Scala collections

❖ Classification of collections

❖ The difference between Iterator and Iterable in Scala and 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 and double-ended queue Deque, Stacks

❖ Sets, Maps and 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 and examples of Spark Split and

  Spark Scala

Introduction to Spark

- ❖ Introduction to Spark and how Spark overcomes the drawbacks of working MapReduce

- ❖ Understanding in-memory MapReduce, interactive operations on MapReduce

- ❖ Spark stack, fine vs. coarse-grained update

- ❖ Spark stack and Spark Hadoop YARN

- ❖ HDFS Revision

- ❖ YARN Revision

- ❖ The overview of Spark and how it is better Hadoop

- ❖ Deploying Spark without Hadoop

- ❖ Spark history server and Cloudera distribution

Spark Basics

- ❖ Spark installation guide and Spark configuration

- ❖ Memory management

- ❖ Executor memory vs. driver memory and working with Spark Shell

- ❖ The concept of resilient distributed datasets (RDD)

- ❖ Learning to do functional programming in Spark and the architecture of Spark

Working with RDDs in Spark

❖ Spark RDD and creating RDDs

❖ RDD partitioning, operations, and transformation in RDD

❖ Deep dive into Spark RDDs

❖ The RDD general operations, a read-only partitioned collection of records, using the concept of RDD for faster and efficient data processing

❖ RDD action for collect, count, collects map, save-as-text-files and pair RDD functions

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 and coarse-grained update and 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 application

❖ Creation of mutable list

❖ Set and set operations, list, tuple, concatenating list

❖ Creating 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 and coalesce and RDD actions

Spark RDD Persistence

- ❖ The execution flow in Spark, understanding the RDD persistence overview

- ❖ Spark execution flow, and Spark terminology

- ❖ Distribution shared memory vs. RDD and RDD limitations

- ❖ Spark shell arguments and Distributed persistence

- ❖ RDD lineage

- ❖ Key-Value pair for sorting implicit conversions like CountByKey, ReduceByKey,

  SortByKey and AggregateByKey

Spark MLlib

- ❖ Introduction to Machine Learning

- ❖ Types of Machine Learning, introduction to MLlib

- ❖ Various ML algorithms supported by MLlib

- ❖ Linear Regression and Logistic Regression

- ❖ Decision Tree and Random Forest

- ❖ K-means clustering techniques, building a Recommendation Engine

Integrating Apache Flume and Apache Kafka

- ❖ Why Kafka and What is Kafka
- ❖ Kafka architecture
- ❖ Kafka workflow
- ❖ Configuring Kafka cluster
- ❖ Basic operations
- ❖ Kafka monitoring tools
- ❖ Integrating Apache Flume and Apache Kafka

Spark Streaming

- ❖ Introduction to Spark Streaming
- ❖ Features of Spark Streaming
- ❖ Spark Streaming workflow
- ❖ Initializing StreamingContext
- ❖ Discretized Streams (DStreams)
- ❖ Input DStreams and Receivers, transformations on Dstreams
- ❖ Output Operations on Dstreams
- ❖ Windowed Operators and why it is useful
- ❖ Important Windowed Operators and Stateful Operators.

Improving Spark Performance

- ❖ Introduction to various variables in Spark like shared variables and 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 structured data processing

- ❖ JSON support in Spark SQL

- ❖ Working with XML data

- ❖ Parquet files

- ❖ Creating Hive context and writing Data Frame to Hive

- ❖ Reading JDBC files and understanding the Data Frames in Spark

- ❖ Creating Data Frames

- ❖ Manual inferring of schema

- ❖ Working with CSV files and reading JDBC tables, Data Frame to JDBC

- ❖ User-defined functions in Spark SQL

- ❖ Shared variables and accumulators

- ❖ Learning to query and transform data in Data Frames

- ❖ How Data Frame provides the benefit of both Spark RDD and Spark SQL and deploying Hive on Spark as the execution engine

Scheduling/Partitioning

- ❖ Learning about the scheduling and partitioning in Spark

- ❖ Hash partition and range partition

- ❖ Scheduling within and around applications

- ❖ Static partitioning and dynamic sharing

- ❖ Fair scheduling and Map partition with index, the Zip, GroupByKey

- ❖ Spark master high availability

- ❖ Standby masters with ZooKeeper

- ❖ Single-node Recovery with Local File System and High Order Functions

Project Works

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. You will understand how to deploy collaborative filtering, clustering, regression, and dimensionality reduction in MLlib. Upon the 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 Dataset

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

Job Assistance Program

Intellipaat is offering job assistance to all the learners who have completed the training. You should get a minimum of 60% marks in the qualifying exam to avail job assistance. Intellipaat has exclusive tie-ups with over 80 MNCs for placements.

Successfully finish the training   Get your resume updated   Start receiving interview calls

Intellipaat Alumni Working in Top Companies

Anthony Crenshaw

Nokia Airscale Integrator (RF) at Nokia

Extensive training
I am glad that I took the Intellipaat Spark training. The trainers offered quality Spark training with real-world examples, and there was extensive interactivity throughout the training that made the Intellipaat training the best according to me

Atyant Jain

Consultant (BI & Advanced Analytics) at Adidas
Real world projects
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 that helped me get hands-on experience in one of the fastest Big Data processing engines. Thank you Intellipaat.
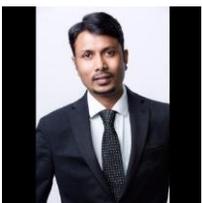
Suman Gajavelly

CTO/Co- Founder Splunk Solutions at bitsIO Inc.
Nice work
I firmly believe that Intellipaat is the perfect place to embark on a great professional career in the technology space. Their Apache Spark and Scala training course was praiseworthy.

More Customer Reviews

Our Clients

+80 Corporates

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. 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 4. 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 5. 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.