# BIG DATA ARCHITECT MASTERS PROGRAM

**IntelliPaat**

## 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, help them choose the right domain, provide extensive hands-on learning experience, rigorously evaluate their 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
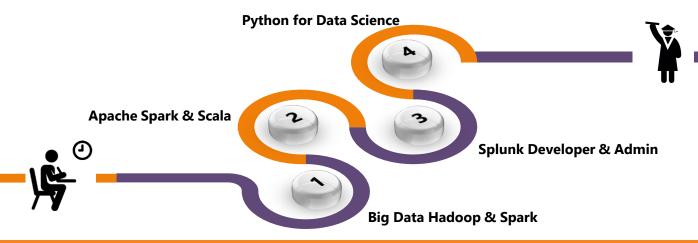
# About The Course

Intellipaat's Big Data Architect master's course will provide you with in-depth knowledge on Big Data platforms like Hadoop, Spark and NoSQL databases, along with a detailed exposure of analytics and ETL by working on tools. This program is specially designed by industry experts, and you will get 12 courses with 31 industry-based projects.

**Instructor Led Training**

173 Hrs of highly interactive instructor led training

**Self-Paced Training**

277 Hrs of Self-Paced sessions with Lifetime access

**Exercise and project work**

384 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.

# Learning Path

**Python for Data Science**

**Apache Spark & Scala**

**Splunk Developer & Admin**

**Big Data Hadoop & Spark**

# Why take this Course?

Big Data is the fastest growing and most promising technology jobs like Big Data Engineer, Big Data Solution Architect are in huge demand. This Big Data Architect course will help you grab the best jobs in this domain This Intellipaat training program has been specifically created to let you masters the Hadoop architect, along with help you gain proficiency in business intelligence domain. Upon completion of the training you will be well-versed in extracting valuable insights from data and convert it into business insights. This way you can apply for the top jobs in the Big Data ecosystem

## Index

### Instructor Led

1. Big Data Hadoop & Spark
2. Apache Spark and Scala
3. Splunk Developer and Admin
4. Python for Data Science
5. Pyspark
6. MongoDB

### Self-Paced Courses

1. Hadoop Testing
2. Apache Storm
3. Apache Kafka
4. Apache Cassandra
5. JAVA
6. LINUX

# Big Data Hadoop & Spark

## Learning Objective

It is a comprehensive Hadoop Big Data training course designed by industry experts considering current industry job requirements to help you learn Big Data Hadoop and Spark modules. This is an industry-recognized Big Data certification training course that is a combination of the training courses in Hadoop developer, Hadoop administrator, Hadoop testing and analytics with Apache Spark. This Cloudera Hadoop and Spark training will prepare you to clear Cloudera CCA175 Big Data certification.

## Topics

- Hadoop Installation and Set-up
- Introduction to Big Data Hadoop. Understanding HDFS & Mapreduce
- Deep Dive in MapReduce
- Introduction to Hive
- Advance Hive & Impala
- Introduction to Pig
- Flume, Sqoop & HBase
- Writing Spark Applications using Scala
- Spark framework
- RDD in Spark
- Data frames and Spark SQL
- Machine Learning using Spark (Mlib)

- Integrating Apache Flume and Apache Kafka
- Spark Streaming
- Hadoop Administration - Multi Node Cluster Setup using Amazon ec2
- Hadoop Administration - Cluster Configuration
- Hadoop Administration - Maintenance, Monitoring and Troubleshooting
- ETL connectivity with Hadoop Ecosystem
- Project Solution Discussion and Cloudera Certification Tips & Tricks

# Apache Spark and Scala

## Learning Objective

It offer you 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.

## Topics

- Introduction of Scala
- Pattern Matching
- Executing the Scala Code
- Classes concept in Scala
- Case classes and pattern matching
- Concepts of traits with example
- Scala java interoperability
- Scala collections
- Mutable collections vs. Immutable collections
- Use case Bobsrockets package

- Introduction to Spark
- Spark Basics
- Working with RDDs in Spark
- Aggregating Data with Pair RDDs
- Writing and Deploying Spark Application
- Parallel Processing
- Spark RDD Persistence
- Spark Streaming & Mlib
- Improving Spark Performance
- Spark SQL and Data Frames
- Scheduling / Partitioning

# Splunk Developer and Admin

## Learning Objective

This is an Intellipaat masters' program in Splunk tool that includes Splunk developer and Splunk administration training. As part of this Splunk course, you will work on searching, sharing, saving Splunk results, creating tags, generating reports and charts, installing and configuring Splunk, monitoring, scaling and indexing large volumes of searches and analyzing it using the Splunk tool.

## Topics

- Splunk Development concepts
- Basic Searching & Using Fields in Searches
- Saving and Scheduling Searches
- Creating Alerts & Scheduled Reports
- Tags and Event Types
- Creating and Using Macros
- Workflow & Splunk Search Commands
- Transforming & Reporting Commands
- Mapping and Single Value Commands
- Splunk Reports & visualizations
- Analyzing, Calculating and Formatting Results
- Correlating Events & Enriching Data with Lookups
- Creating Reports and Dashboards
- Getting started with Parsing
- Using Pivot

- Common Information Model (CIM) Add-On
- Overview of Splunk and its Installation
- Distributed Management Console
- Introduction to Splunk App
- Splunk indexes and users
- Splunk configuration files
- Splunk Deployment Management
- Splunk Indexes
- User roles and authentication
- Splunk Administration Environment
- Basic Production Environment
- Splunk Search Engine
- Various Splunk Input Methods
- Splunk User & Index Management
- Machine Data Parsing
- Search Scaling and Monitoring
- Splunk Cluster implementation

# Python for Data Science

## Learning Objective

Intellipaat Python for Data Science training helps you learn the top programming language for the domain of Data Science. You will master the technique of how Python is deployed for Data Science, work with Pandas library for Data Science, do data munging and data cleaning, advanced numeric analysis and more through real-world hands-on projects and case studies.

## Topics

❖ Introduction to Data Science

❖ Introduction to Python

❖ Python basic constructs

❖ OOPs in Python

❖ NumPy for mathematical computing

❖ SciPy for scientific computing

❖ Data manipulation

❖ Data visualization with Matplotlib

❖ Machine Learning using Python

❖ Supervised learning

❖ Unsupervised Learning

❖ Python integration with Hadoop and Spark

# Self-Paced Courses

| | |
|---|---|
| **Hadoop Testing** | This Hadoop Testing course will provide you with the right skills to detect, analyze and rectify errors in Hadoop framework. You will be trained in the Hadoop software, architecture, MapReduce, HDFS, and various components like Pig, Hive, Sqoop, Flume and Oozie. With this Hadoop testing training you will be fully equipped in various test case scenarios, Proof of Concepts implementation and real world scenarios. |
| **Apache Storm** | Intellipaat Apache Storm Certification course lets you master distributed, stream processing engine Storm. You will learn about the deployment and development of Apache Storm applications in real world for handling big data and implement various analytical tools for powerful enterprise-grade solutions. |
| **Apache Kafka** | Our Apache Kafka training course will give you hands-on experience to master the real-time stream processing platform. Some of the topics included in this online training course are the Kafka API, creating Kafka clusters, integration of Kafka with the Big Data Hadoop ecosystem along with Spark, Storm and Maven integration. |
| **Cassandra** | This Cassandra course will provide you extensive knowledge on Cassandra concepts, high-scalable data models and the Cassandra architecture which will enable you to build applications for big data. You will learn Cassandra configuration, installation, architecture, data modelling, Hadoop integration. Work on real life industry projects. |
| **Java** | This comprehensive course lets you master Java programming language. We provide the best online training classes to help you learn OOP concepts, J2EE, core and advanced Java, JDBC, Objects & Classes. You will work on real world industry projects. |
| **Linux** | This is an in-depth Training Course that gives you all the skills needed for working as a Linux Administrator. You will learn about the Red Hat System, installation, managing the boot processes, performing various operations, understanding Linux Kernel, testing and debugging. |
| **MongoDB** | This is a very extensive course in MongoDB that is one of the most widely used NoSQL tools in the Big Data domain. Some of the topics that are included in this MongoDB Training  are installation of MongoDB,  JSON, Data Modeling and Schema Design. You will gain enough expertise in the framework of Data Monitoring, Indexing, and Aggregation. |

# Project Works

| Course | Industry/Domain | Project |
|---|---|---|
| **Big Data Hadoop & Spark** | General | Working with Mapreduce, Hive, Sqoop |
| | Media | Work on MovieLens data for findin top movies |
| | Banking | Hadoop YARN project - End to End PoC |
| | Banking | Table Partitioning in Hive |
| | Sociall Network | Connecting Pentaho with Hadoop Ecosystem |
| | General | Multi Node Cluster setup |
| | General | Hadoop Testing using MRUnit |
| | Internet Services | Hadoop Weblog Analytics |
| | General | Hadoop Maintenance |
| | Sociall Network | Twitter Sentimental Analysis |
| | Sports | Analysing IPL T20 Cricket |
| **Apache Spark and Scala** | Entertainment | Movie Recommendation |
| | Social Media | Twitter API integration for Tweet Analysis |
| | Internet Services | Data Exploration using Spark SQL – Wikipedia Data set |
| **Splunk Developer and Admin** | General | Creating an Employee Database of a Company |
| | E-Commerce | Building an Organizational Dashboard on Splunk |
| | General | Field Extraction in Splunk |
| **Python for Data Science** | General | Analysing the naming pattern using Python |
| | General | Python Web Scraping for Data Science |
| | Telecommunications | Predicting customer churn in Telecom company |
| | Internet Services | Server logs/ Firewall logs |
| **+10 additional projects as part of self-paced courses** | | |

# 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.

## 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.