



Big Data Architect Master's Course

In Collaboration with **IBM**

Table of Contents

1. About the Program
2. Collaborating with IBM
3. About Intellipaap
4. Key Features
5. Career Support
6. Why take up this course?
7. Who should take up this course?
8. Learning Path
9. Course Advisors
10. Program Curriculum
11. Self-paced Courses
12. Project Work
13. Certification
14. Intellipaap Success Stories
15. Contact Us



About the Program

Our Big Data Architect master's course lets you gain proficiency in Big Data. You will work on real-world projects in Hadoop Development, Hadoop Administration, Hadoop Analysis, Hadoop Testing, Spark, Python, Splunk Developer and Admin, Apache Storm, NoSQL databases, and more. In this program, you will cover 13 courses and 33 industry-based projects. As part of this online classroom training, you will receive four additional self-paced courses co-created with IBM, namely, Spark Fundamentals I and II, Spark MLlib, and Python for Data Science.

IBM

Collaborating with IBM

IBM is one of the leading innovators and the biggest player in creating innovative tools for Big Data Analytics. Top subject matter experts from IBM will share knowledge in the domains of Analytics and Big Data through this training program, which will help you gain the breadth of knowledge and industry experience.

Benefits for students from IBM

- Industry-recognized IBM certificate
- Access to IBM Watson for hands-on training and practice
- Industry in-line case studies and project work



At Intellipaat, we are in a constant endeavor of changing the education culture to be more of an open-source collaborative environment.

About Intellipaat

Intellipaat is one of the leading e-learning training providers with more than 600,000 learners across 55+ countries. We are on a mission to democratize education as we believe that everyone has the right to quality education.

Our courses are delivered by subject matter experts from top MNCs, and our world-class pedagogy enables learners to quickly learn difficult topics in no time. Our 24/7 technical support and career services will help them jump-start their careers in their dream companies.

Key Features



**205 HRS OF INSTRUCTOR-LED
TRAINING**



**277 HRS OF SELF-PACED
TRAINING**



**384 HRS OF REAL-TIME
PROJECT WORK**



LIFETIME ACCESS



24/7 TECHNICAL SUPPORT



**INDUSTRY-RECOGNIZED
CERTIFICATION**



**JOB ASSISTANCE THROUGH
80+ CORPORATE TIE-UPS**



FLEXIBLE SCHEDULING

Career Support



SESSIONS WITH INDUSTRY MENTORS

Attend sessions from top industry experts and get guidance on how to boost your career growth



MOCK INTERVIEWS

Mock interviews to make you prepare for cracking interviews by top employers



GUARANTEED INTERVIEWS & JOB SUPPORT

Get interviewed by our 400+ hiring partners



RESUME PREPARATION

Get assistance in creating a world-class resume from our career services team



Why take up this course?

- Global Hadoop market to reach US\$84.6 billion in 2 years – Allied Market Research
- The number of jobs for all US-based data professionals will increase by 2.7 million per year – IBM
- A Hadoop Administrator in the United States can get a salary of US\$123,000 – Indeed

Big Data is the fastest growing and the most promising technology that aids profiles, such as Big Data Engineer and Big Data Solutions Architect, that are in huge demand. This Big Data Architect master's course will help you grab the best jobs in this domain.

Who should take up this training?

- Data Science and Big Data Professionals and Software Developers
- Business Intelligence Professionals, Information Architects, and Project Managers
- Those who aspire to be a Big Data Architect

Courses offered

INSTRUCTOR-LED TRAINING COURSES

- Big Data Hadoop and Spark
- Apache Spark and Scala
- Splunk Developer and Admin
- Python for Data Science
- PySpark Training
- MongoDB
- AWS Big Data

SELF-PACED LEARNING COURSES

- Hadoop Testing
- Apache Storm
- Apache Kafka
- Apache Cassandra
- Java
- Linux

Course Advisors



Suresh Paritala

Solutions Architect at Microsoft, Texas

A Senior Software Architect at Microsoft, Texas, who has previously worked with IBM Corporation, Suresh Paritala has worked on Big Data, Data Science, Advanced Analytics, Internet of Things, and Azure, along with AI domains like Machine Learning and Deep Learning. He has successfully implemented high-impact projects in major corporations around the world.



David Callaghan

Big Data Strategist and Solutions Architect, Perficient, USA

An experienced Blockchain professional, who has been bringing integrated Blockchain, particularly Hyperledger and Ethereum, and Big Data solutions to the cloud, David Callaghan has previously worked on Hadoop, AWS Cloud, Big Data, and Pentaho projects that have had a major impact on the revenues of marquee brands around the world.



Samanth Reddy

Data Scientist at ASCAP

A renowned Data Scientist who has worked with Google and is currently working at ASCAP, Samanth Reddy has a proven ability to develop Data Science strategies that have a high impact on the revenues of various organizations. He comes with strong Data Science expertise and has created decisive Data Science strategies for Fortune 500 corporations.

Program Curriculum

BIG DATA HADOOP & SPARK

Learning Objective

IntelliPaat Big Data Hadoop training program helps you master Big Data Hadoop and Spark to get ready for the Cloudera CCA Spark and Hadoop Developer Certification (CCA175) exam as well as master Hadoop Administration with 14 real-time industry-oriented case-study projects. In this Big Data course, you will master MapReduce, Hive, Pig, Sqoop, Oozie and Flume and work with Amazon EC2 for cluster setup, Spark framework and RDD, Scala and Spark SQL, Machine Learning using Spark, Spark Streaming, etc.

Topics:

- Hadoop Installation and Setup
- Introduction to Big Data Hadoop and Understanding HDFS and MapReduce
- Deep Dive into MapReduce
- Introduction to Hive
- Advanced Hive and Impala
- Introduction to Pig
- Flume, Sqoop, and HBase
- Writing Spark Applications Using Scala
- Introduction to Spark
- Spark Basics
- Working with RDDs in Spark
- Aggregating Data with Pair RDDs
- Writing and Deploying Spark Applications

- Project Solution Discussion and Cloudera Certification Tips and Tricks
- Parallel Processing
- Spark RDD Persistence
- Spark MLlib
- Integrating Apache Flume and Apache Kafka
- Spark Streaming
- Improving Spark Performance
- Spark SQL and Data Frames
- Scheduling/Partitioning

Self-paced Course Content:

- Hadoop Administration – Multi-node Cluster Setup Using Amazon EC2
- Hadoop Administration – Cluster Configuration
- Hadoop Administration – Maintenance, Monitoring and Troubleshooting
- ETL Connectivity with Hadoop Ecosystem (Self-Paced)
- Hadoop Application Testing
- Roles and Responsibilities of Hadoop Testing Professional
- Framework Called MRUnit for Testing of MapReduce Programs
- Unit Testing
- Test Execution
- Test Plan Strategy and Writing Test Cases for Testing Hadoop Application

APACHE SPARK & SCALA

Learning Objective

Intellipaate Spark training lets you master real-time data processing using Spark streaming, Spark SQL, Spark RDD and Spark Machine Learning libraries (Spark MLlib). You will learn Spark and Scala programming, as well as work on three real-life use

cases in this Spark and Scala course.

Topics:

- Introduction to Scala
- Pattern Matching
- Executing the Scala Code
- Classes Concept in Scala
- Case Classes and Pattern Matching
- Concept 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 Paired RDDs
- Writing and Deploying Spark Applications
- Parallel Processing
- Spark RDD Persistence
- Spark MLlib
- Integrating Apache Flume and Apache Kafka
- Spark Streaming
- Improving Spark Performance
- Spark SQL and DataFrames
- Scheduling/Partitioning

SPLUNK DEVELOPER & ADMIN

Learning Objective

The IntelliPaat Splunk certification training includes the complete aspects of Splunk developer and Splunk administration. This Splunk course also includes various aspects of Splunk installation, configuration, Splunk Syslog, Syslog Server, log analysis, Splunk dashboard, installation, configuration of Splunk, deploying Splunk search, monitor, index, report and analysis.

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 Commands
- Reporting Commands
- Mapping and Single-value Commands
- Splunk Reports and 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
- Splunk Installation
- Splunk Installation in Linux
- Distributed Management Console
- Introduction to the 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 and Index Management
- Machine Data Parsing
- Search Scaling and Monitoring
- Splunk Cluster Implementation

PYTHON FOR DATA SCIENCE

Learning Objective

The Data Science with Python course enables you to master Data Science Analytics using Python. You will work on various Python libraries such as SciPy, NumPy, Matplotlib, Lambda function, etc. You will master Data Science Analytics skills through real-world projects covering multiple domains such as retail, e-commerce, finance, etc.

Topics:

- Introduction to Data Science Using Python
- Python Basic Constructs
- Maths for DS: Statistics and Probability
- 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 Spark (Self-paced)
- Dimensionality Reduction
- Time Series Forecasting

PYSPARK

Learning Objective

Intellipaats PySpark course is designed to help you understand the PySpark concept and develop custom, feature-rich applications using Python and Spark. Our PySpark training

courses are conducted online by leading PySpark experts working in top MNCs. During this PySpark course, you will gain in-depth knowledge of Apache Spark and related ecosystems, including Spark Framework, PySpark SQL, PySpark Streaming, and more. In addition, you can work in a virtual lab and run real-time projects to get hands-on experience with PySpark.

Topics:

- Introduction to the Basics of Python
- Sequence and File Operations
- Functions, Sorting, Errors and Exception, Regular Expressions, and Packages
- Python: An OOP Implementation
- Debugging and Databases
- Introduction to Big Data and Apache Spark
- Python for Spark
- Python for Spark: The Functional and Object-oriented Model
- Apache Spark Framework and RDDs
- PySpark SQL and DataFrames
- Apache Kafka and Flume
- PySpark Streaming
- Introduction to PySpark Machine Learning

MONGODB DEVELOPER & ADMIN

Learning Objective

Our MongoDB certification training course will help you master the NoSQL database. We provide the best online classes to help you learn MongoDB installation, data modeling, schema design, data indexing, monitoring, and aggregation. The course also offers opportunities to work on real-world projects.

Topics:

- Introduction to NoSQL and MongoDB
- MongoDB Installation
- Importance of NoSQL
- CRUD Operations
- Data Modeling and Schema Design
- Data Management and Administration
- Data Indexing and Aggregation
- MongoDB Security
- Working with Unstructured Data

AWS BIG DATA

Learning Objective

This AWS Big Data certification course will help you gain in-depth knowledge of AWS Big Data concepts, such as AWS IoT (Internet of Things), Kinesis, Amazon DynamoDB, Amazon Machine Learning (AML), data analysis, data processing technologies, data visualization, and more. Through this AWS Big Data training, you will be able to clear the AWS Certified Data Analytics - Specialty exam, DAS-C01.

Topics:

- Introduction to Big Data and Data Collection
- Introduction to Cloud Computing and AWS
- Elastic Compute and Storage Volumes
- Virtual Private Cloud
- Storage - Simple Storage Service (S3)
- Databases and In-memory Data Stores
- Data Storage
- Data Processing

- Data Analysis
- Data Visualization and Data Security

Self-paced Courses

As part of this online classroom training, you will receive six additional self-paced courses co-created with IBM, namely, Hadoop Testing, Apache Storm, Apache Kafka, Apache Cassandra, Java, and Linux. Moreover, you will also get exclusive access to IBM Watson Cloud Lab for the Chatbots course.

Project Work

Big Data Hadoop Projects

- Working with MapReduce, Hive, and Sqoop
- Working on MovieLens Data for Finding the Top Movies
- Hadoop YARN Project: End-to-End PoC
- Table Partitioning in Hive
- Connecting Pentaho with the Hadoop Ecosystem
- Multi-node Cluster Setup
- Hadoop Testing Using MRUnit
- Hadoop Web Log Analytics
- Hadoop Maintenance
- Twitter Sentiment Analysis
- Analyzing IPL T20 Cricket
- Movie Recommendation
- Twitter API Integration for Tweet Analysis
- D Data Exploration Using Spark SQL – Wikipedia Dataset

Apache Spark & Scala

- Movie Recommendation
- Twitter API Integration for Tweet Analysis
- Data Exploration Using Spark SQL – Wikipedia Dataset

Splunk

- Creating an Employee Database of a Company
- Building an Organizational Dashboard with Splunk
- Field Extraction in Splunk

Python for Data Science

- Analyzing the Trends of COVID-19 with Python
- Analyzing the Naming Trends Using Python
- Performing Analysis on Customer Churn Dataset
- Netflix Recommendation System
- Python Web Scraping for Data Science
- OOPS in Python
- Working With NumPy
- Visualizing and Analyzing the Customer Churn Dataset Using Python
- Building Models with the Help of Machine Learning Algorithms

MongoDB

- Working with the MongoDB Java Driver

AWS Big Data

- Integration of Big Data with AWS
- Big Data Analysis

Certification

After the completion of the course, you will get certificates from IBM and IntelliPaat.



CERTIFICATE OF COMPLETION

This Certificate Is Presented To

Your Name

Who has successfully completed all the requirements stipulated by IntelliPaat for

Course Name

to achieve professional excellence

Issue Date: January 2, 2020

Mrs. Shilpi Jain
Director,
IntelliPaat Software Solutions Pvt Ltd.

VERIFIED
CERTIFICATE

Certificate ID 31679-119525-18232

Success Stories



Vishal Pentakota

The best part of this online course is the series of hands-on demonstrations the trainer performed. Not only did he explain each concept theoretically, but he also implemented all those concepts practically. Great job! A must go for beginners.



Shreyashkumar Limbhetwala

I want to talk about the rich LMS that Intellipaat's Data Science training offered. The extensive set of PPTs, PDFs, and other related course material were of the highest quality, and due to this, my learning with Intellipaat was excellent. I could clear the certification in the first attempt.



Giri Karnal

I had taken up the Big Data Architect master's program, which is a combo of so many courses. Since there are so many technologies involved in the program, getting your query resolved at the right time becomes the most important aspect. But with Intellipaat, there was no such problem as all my queries were resolved in less than 24 hours.



Sharath Reddy Yellapati

The course material was well organized. The trainer explained each module to me from the basics. All my queries were addressed very clearly. The trainer also made me realize how important this course is for beginners in the IT stream.

CONTACT US

INTELLIPAAT SOFTWARE SOLUTIONS PVT. LTD.

Bangalore

AMR Tech Park 3, Ground Floor, Tower B,
Hongasandra Village, Bommanahalli,
Hosur Road, Bangalore – 560068

USA

1219 E. Hillsdale Blvd. Suite 205,
Foster City, CA 94404

If you have any further queries or just want to have a conversation with us, then do call us.

IND: +91-7022374614 | US: 1-800-216-8930