



CERTIFICATE PROGRAM

Big Data with Spark and Scala

Online Self Paced Course | 30+ Hours of Training

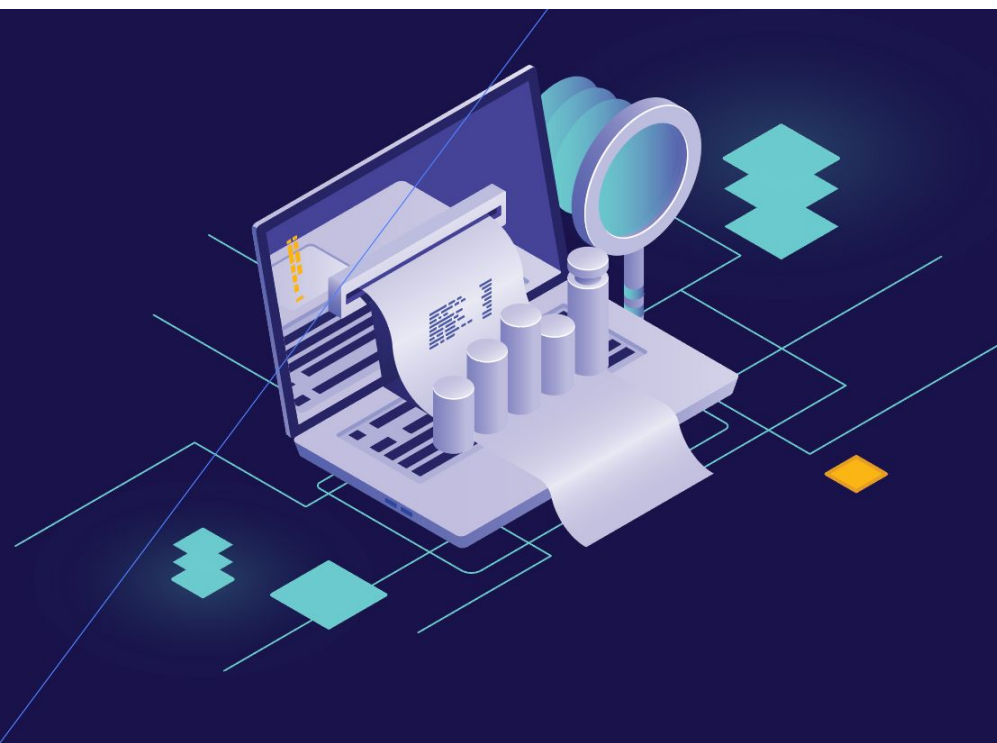


CloudxLab & Course

At Cloudxlab, we are building one of the best gamified learning environments to make technology learning fun and for life. More than 50,000 users across the world have been benefited by our signature courses on Machine Learning and Big Data. Our vision is to upskill people on high-end technologies like Deep Learning, Machine Learning, Big Data and make them employable.

As humans, we are immersed in data in our everyday lives. As per IBM, the data doubles every two years on this planet. The value that data holds can only be understood when we can start to identify patterns and trends in the data. Normal computing principles do not work when data becomes huge.

There is massive growth in the big data space, and job opportunities are skyrocketing, making this the perfect time to launch your career in this space. In this course, you will learn Spark to drive better business decisions and solve real-world problems.



Sandeep Giri

Founder at CloudxLab

Why CloudxLab



Earn a Verified Certificate from CloudxLab



Learn Big Data with Spark from industry experts and become expert in Data Science domain



Online cloud lab for hands-on for real-world experience



Best-in-class support Throughout your learning journey



Lifetime course access



Work on real-world projects.



Interact with the international community of peers via the discussion forum.

Course Creators



Sandeep Giri

Founder at CloudxLab

Past: Amazon, InMobi, D.E.Shaw

Course Developer

[Know More](#)



Abhinav Singh

Co-Founder at CloudxLab

Past: Byjus

Course Developer

[Know More](#)



Jatin Shah

Ex-LinkedIn, Yahoo,

Yale CS Ph.D. IIT-B

Course Advisor

[Know More](#)



Praveen Pavithran

Co-Founder at Yatis

Past: YourCabs, Cypress Semiconductor

Course Advisor

[Know More](#)

Course Curriculum

1. Introduction

- Apache Spark ecosystem walkthrough
- Spark Introduction - Why Spark?
- Quiz

2. Scala Basics

- Scala - Quick Introduction - Access Scala on CloudxLab
- Scala - Quick Introduction - Variables and Methods
- Getting Started: Interactive, Compilation, SBT
- Types, Variables & Values
- Functions
- Collections
- Classes
- Parameters
- More Features
- Quiz and Assessment

3. Spark Basics

- Apache Spark ecosystem walkthrough
- Spark Introduction - Why Spark?
- Using the Spark Shell on CloudxLab
- Example 1 - Performing Word Count
- Understanding Spark Cluster Modes on YARN
- RDDs (Resilient Distributed Datasets)
- General RDD Operations: Transformations & Actions
- RDD lineage
- RDD Persistence Overview
- Distributed Persistence.

Course Curriculum ---

4. Writing and Deploying Spark Applications

- Creating the SparkContext
- Building a Spark Application (Scala, Java, Python)
- The Spark Application Web UI
- Configuring Spark Properties
- Running Spark on Cluster
- RDD Partitions
- Executing Parallel Operations
- Stages and Tasks

5. Common Patterns in Spark Data Processing

- Common Spark Use Cases
- Example 1 - Data Cleaning (Movielens)
- Example 2 - Understanding Spark Streaming
- Understanding Kafka
- Example 3 - Spark Streaming from Kafka
- Iterative Algorithms in Spark
- Project: Real-time analytics of orders in an e-commerce company

Course Curriculum

6. Data Formats and Management

- InputFormat and InputSplit
- JSON
- XML
- AVRO
- How to store many small files - SequenceFile?
- Parquet
- Protocol Buffers
- Comparing Compressions
- Understanding Row Oriented and Column Oriented Formats - RCFile?

7. DataFrames and Spark SQL

- Spark SQL - Introduction
- Spark SQL - Dataframe Introduction
- Transforming and Querying DataFrames
- Saving DataFrames
- DataFrames and RDDs
- Comparing Spark SQL, Impala, and Hive-on-Spark

8. Machine Learning with Spark

- Machine Learning Introduction
- Applications Of Machine Learning
- MLlib Example: k-means
- SparkR Example

Projects

1. Spark MLlib

Generate movie recommendations using Spark MLlib

2. Spark GraphX

Derive the importance of various handles at Twitter using Spark GraphX

3. Churn the logs

Churn the logs of NASA Kennedy Space Center WWW server using Spark to find out useful business and devops metrics

4. Spark application

Write end-to-end Spark application starting from writing code on your local machine to deploying to the cluster

5. Analytics Dashboard

Real-time analytics dashboard for an e-commerce company using Apache Spark, Kafka, Spark Streaming, Node.js, Socket.IO and Highcharts

Course Details and Fees —

Please find more information about the course and fees here:

<https://cloudxlab.com/course/4/big-data-with-spark>

Mode of Learning —

Online Self-Paced Learning

Our Esteemed Customers —

simplilearn

greatlearning

INSOFE
Inspire...Educate...Transform.

Berkeley
UNIVERSITY OF CALIFORNIA

Udemy

Tech
Mahindra



Cornell University

HARVARD
UNIVERSITY

Mit
Massachusetts
Institute of
Technology

Carnegie
Mellon
University

W
UNIVERSITY of WASHINGTON

For Further Details —

Contact us at [+080-4920-2224](tel:+080-4920-2224) or [+1 412-568-3901](tel:+1412-568-3901) or contact:



Aswath Madhu
Program Director

programs@cloudxlab.com



Prakhar Katiyar
Chief Admissions Counsellor

admissions@cloudxlab.com

For Business —

For corporate training and bulk enrollments, write to us at reachus@cloudxlab.com

Headquarters - United States

2035, Sunset Lake Road Suite B-2, 19702
Newark, New Castle
Delaware, United States

R&D Center - India

Issimo Technology Private Limited
#215, Arcade, Brigade Metropolis,
Mahadevpura, Bangalore, India - 560 048