



**CERTIFICATE PROGRAM**

# **Big Data with Hadoop**

**Online Self Paced Course | 30+ Hours of Training**



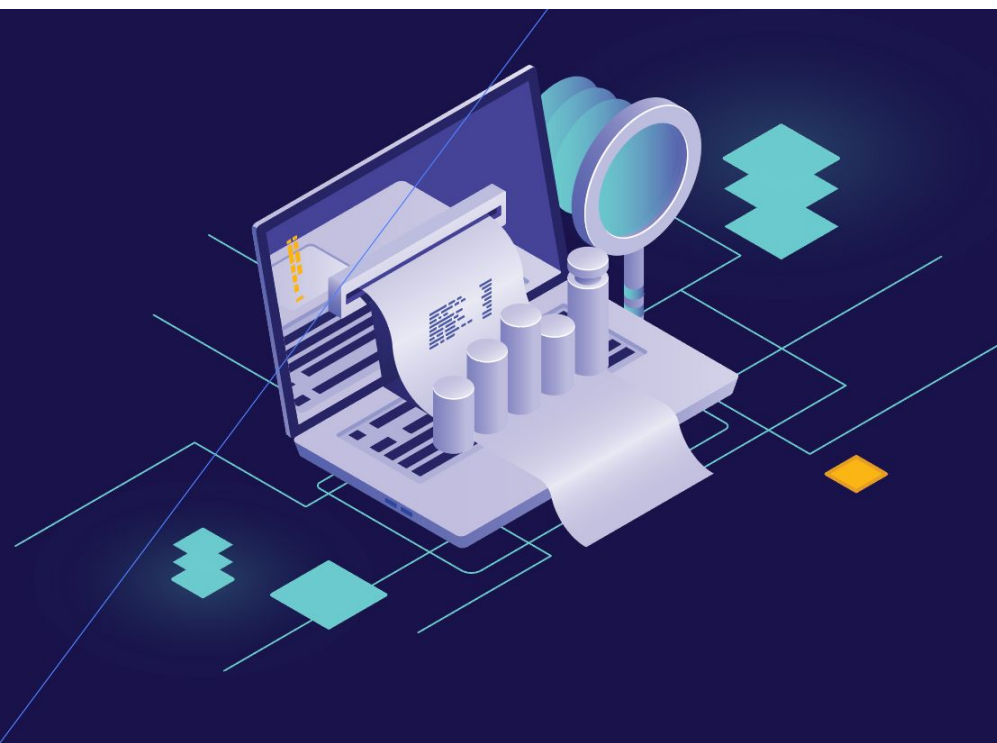
# CloudxLab & Course

---

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 Hadoop 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 Hadoop 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

- Big Data Introduction
- Distributed systems
- Big Data Use Cases
- Various Solutions
- Overview of Hadoop Ecosystem
- Spark Ecosystem Walkthrough
- Quiz

## 2. Foundation & Environment

- Understanding the Cloudxlab
- Cloudxlab Hands-on
- Hadoop & Spark Hands-on
- Quiz and Assessment
- Basics of Linux - Quick Hands-on
- Understanding Regular Expressions
- Quiz and Assessment
- Setting up VM (optional)

# Course Curriculum

---

## 3. Zookeeper

- ZooKeeper - Race Condition
- ZooKeeper - Deadlock
- Hands-On
- Quiz & Assessment
- How does election happen - Paxos Algorithm?
- Use cases
- When not to use
- Quiz & Assessment

## 4. HDFS

- Why HDFS or Why not existing file systems?
- HDFS - NameNode & DataNodes
- Quiz
- Advance HDFS Concepts (HA, Federation)
- Quiz
- Hands-on with HDFS (Upload, Download, SetRep)
- Quiz & Assessment
- Data Locality (Rack Awareness)

## 5. YARN

- YARN - Why not existing tools?
- YARN - Evolution from MapReduce 1.0
- Resource Management: YARN Architecture
- Advance Concepts - Speculative Execution
- Quiz

# Course Curriculum

---

## 6. MapReduce Basics

- MapReduce - Understanding Sorting
- MapReduce - Overview & Quiz
- Example 0 - Word Frequency Problem - Without MR
- Example 1 - Only Mapper - Image Resizing
- Example 2 - Word Frequency Problem
- Example 3 - Temperature Problem
- Example 4 - Multiple Reducer
- Example 5 - Java MapReduce Walkthrough & Quiz

## 7. Map Reduce Advanced

- Writing MapReduce Code Using Java
- Building MapReduce project using Apache Ant
- Concept - Associative & Commutative
- Quiz
- Example 8 - Combiner
- Example 9 - Hadoop Streaming
- Example 10 - Adv. Problem Solving - Anagrams
- Example 11 - Adv. Problem Solving - Same DNA
- Example 12 - Adv. Problem Solving - Similar DNA
- Example 12 - Joins - Voting
- Limitations of MapReduce
- Quiz

# Course Curriculum

---

## 8. Analyzing Data with Pig

- Pig - Introduction
- Pig - Modes
- Getting Started
- Example - NYSE Stock Exchange
- Concept - Lazy Evaluation

## 9. Processing Data with Hive

- Hive - Introduction
- Hive - Data Types
- Getting Started
- Loading Data in Hive (Tables)
- Example: Movielens Data Processing
- Advance Concepts: Views
- Connecting Tableau and HiveServer 2
- Connecting Microsoft Excel and HiveServer 2
- Project: Sentiment Analysis of Twitter Data
- Advanced - Partition Tables
- Understanding HCatalog & Impala
- Quiz



# Course Curriculum

---

## 10. NoSQL and HBase

- NoSQL - Scaling Out / Up
- NoSQL - ACID Properties and RDBMS Story
- CAP Theorem
- HBase Architecture - Region Servers etc
- Hbase Data Model - Column Family Orientedness
- Getting Started - Create table, Adding Data
- Adv Example - Google Links Storage
- Concept - Bloom Filter
- Comparison of NOSQL Databases
- Quiz

## 11. Importing Data with Sqoop and Flume, Oozie

- Sqoop - Introduction
- Sqoop Import - MySQL to HDFS
- Exporting to MySQL from HDFS
- Concept - Unbounding Dataset Processing or Stream Processing
- Flume Overview: Agents - Source, Sink, Channel
- Example 1 - Data from Local network service into HDFS
- Example 2 - Extracting Twitter Data
- Quiz
- Example 3 - Creating workflow with Oozie

# Projects

---

## **1. Sentiment analysis**

Sentiment analysis of "Iron Man 3" movie using Hive and visualizing the sentiment data using BI tools such as Tableau

## **2. Process the NSE**

Process the NSE (National Stock Exchange) data using Hive for various insights

## **3. MovieLens Project**

Analyze MovieLens data using Hive

## Course Details and Fees —

Please find more information about the course and fees here:

<https://cloudxlab.com/course/3/big-data-with-hadoop>

## 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](mailto:programs@cloudxlab.com)



**Prakhar Katiyar**  
Chief Admissions Counsellor

[admissions@cloudxlab.com](mailto:admissions@cloudxlab.com)

## For Business —

For corporate training and bulk enrollments, write to us at [reachus@cloudxlab.com](mailto: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