



CERTIFICATION PROGRAM

Data Engineering with Hadoop and Spark

3 Months | Online Self-paced Course



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 benefitted 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 every-day 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 and Spark to drive better business decisions and solve real-world problems.



Sandeep Giri

Founder at CloudXLab

E&ICT Academy, IIT Roorkee

E&ICT (Electronics & Information & Communications Technology) Academy, IIT Roorkee provides training programs with an emphasis on hands-on learning in basic/advanced topics and emerging technologies. The project is sponsored by the Ministry of Electronics and Information Technology, Govt. of India. E&ICT Academy courses are recognized by All India Council for Technical Education (AICTE) at par with QIP for recognition/credits.

The programs are conducted by well-known industry partners, researches and experts from leading academic and renowned R&D organizations. For this, the Academy has signed MoUs with Industry/ R&D partners in different domains, who collaborate and work with them in conducting the training programs. Academy also facilitates the interaction between beneficiaries and industry experts to enable collaboration and finding opportunities for parent institutions.



Sanjeev Manhas

Associate Professor IIT Roorkee

Why CloudxLab and E&ICT Academy, IIT Roorkee



Earn a certificate from E&ICT Academy, IIT Roorkee.



Learn Big Data from industry experts and become expert in Big Data domain.



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



Work on real-world projects.



The course is compatible with these certifications - CCP Data Engineer, CCA Spark, and Hadoop Developer, HDP Certified Developer: Spark



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



Lifetime course access.



Best-in-class support throughout your learning journey.

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 >](#)



Sanjeev Manhas

Associate Professor,
IIT Roorkee

Course Advisor
[Know more >](#)



R. Balasubramanian

Professor,
IIT Roorkee

Course Developer
[Know more >](#)



Partha Pratim Roy

Assistant Professor,
IIT Roorkee

Course Developer
[Know more >](#)

Course Curriculum

1 Big Data with Hadoop

1. Introduction to Big Data
2. Foundation & Environment
3. Zookeeper
4. HDFS
5. YARN
6. MapReduce Basics
7. MapReduce Advanced
8. Analyzing Data with Pig
9. Processing Data with Hive
10. NoSQL and HBase
11. Importing Data with Sqoop and Flume, Oozie

2 Big Data with Spark

1. Scala Basics
2. Spark Basics
3. Writing and Deploying Spark Applications
4. Common Patterns in Spark Data Processing
5. Data Formats & Management
6. DataFrames and Spark SQL
7. Machine Learning with Spark

Projects

- 1** Sentiment analysis of "Iron Man 3" movie using Hive and visualizing the sentiment data using BI tools such as Tableau
- 2** Process the NSE (National Stock Exchange) data using Hive for various insights
- 3** Derive the importance of various handles at Twitter using Spark GraphX
- 4** Build a real-time analytics dashboard for an e-commerce company using Apache Spark, Kafka, Spark Streaming, Node.js, Socket.IO and Highcharts
- 5** Write end-to-end Spark application starting from writing code on your local machine to deploying on the cluster
- 6** Analyze MovieLens data using Hive
- 7** Churn the logs of NASA Kennedy Space Center WWW server using Spark to find out useful business and DevOps metrics
- 8** Generate movie recommendations using Spark MLlib

Course Details and Fees —

Please find more information about the course and fees here:

<https://cloudxlab.com/course/67/big-data-engineering-with-hadoop-and-spark>

Our Esteemed Customers —



Cornell University



For Further Details —

Contact us at **+080-4920-2224** or **+1 412-568-3901** 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