CLOUD x LAB

eict
IIT Roorkee

CERTIFICATE PROGRAM

Data Science Specialization

220+ Hours of Training | 26+ Projects | Online 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 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.

Every domain of computing such as data analysis, software engineering, and artificial intelligence is going to be impacted by Data Science. Therefore,every engineer, researcher, manager or scientist would be expected to know Data Science.

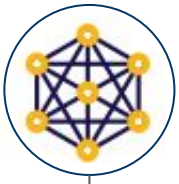
So naturally, you are excited about Data Science and would love to dive into it. This specialization is designed for those who want to gain hands-on experience in solving real-life problems using big data, machine learning and deep learning. After finishing this specialization, you will find creative ways to apply your learning to your work like building a robot which can recognize faces or change the path after discovering obstacles on the path.

Sandeep Giri

Founder at CloudxLab

Why CloudxLab

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



Learn Data Science 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

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

Abhinav Singh

Co-Founder at CloudxLab

Past: Byjus

Course Developer

Know More

Course Curriculum

Course 1: Python For Machine Learning

- Introduction to Linux
- Introduction to Python
- Hands-on using Jupyter on CloudxLab
- Overview of Linear Algebra
- Introduction to NumPy & Pandas

Course 2: Big Data with Hadoop

1. Introduction

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

2. Foundation & Environment

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

Course Curriculum

Course 2: Big Data with Hadoop

3. Zookeeper

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

4. HDFS

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

Course Curriculum

Course 2: Big Data with Hadoop

6. MapReduce Basics

- MapReduce - Understanding Sorting and Overview
- 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

7. Mapreduce Advanced

- Writing MapReduce Code Using Java
- Building MapReduce project using Apache Ant
- Concept - Associative & Commutative
- 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

8. Analyzing Data with Pig

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

Course Curriculum

Course 2: Big Data with Hadoop

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

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

Course Curriculum

Course 2: Big Data with Hadoop

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
- Example 3 - Creating workflow with Oozie

Course Curriculum

Course 3: Big Data with Spark

1. Introduction

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

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

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

Course 3: Big Data with Spark

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

Course 3: Big Data with Spark

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
- MILib Example: k-means
- SparkR Example

Course Curriculum

Course 4: Machine Learning

1. Introduction to Statistic

Statistical Inference, Types of Variables, Probability Distribution, Normality, Measures of Central Tendencies, Normal Distribution

2. Machine Learning Applications & Landscape

Introduction to Machine Learning, Machine Learning Application, Introduction to AI, Different types of Machine Learning - Supervised, Unsupervised, Reinforcement

3. Building end-to-end Machine Learning Project

Machine Learning Projects Checklist, Frame the problem and look at the big picture, Get the data, Explore the data to gain insights, Prepare the data for Machine Learning algorithms, Explore many different models and short-list the best ones, Fine-tune model, Present the solution, Launch, monitor, and maintain the system

4. Classifications

Training a Binary classification, Performance Measures, Confusion Matrix, Precision and Recall, Precision/Recall Tradeoff, The ROC Curve, Multiclass Classification, Multilabel Classification, Multioutput Classification

5. Training Models

Linear Regression, Gradient Descent, Polynomial Regression, Learning Curves, Regularized Linear Models, Logistic Regression

6. Support Vector Machines

Linear SVM Classification, Nonlinear SVM Classification, SVM Regression

Course Curriculum

Course 4: Machine Learning

7. Decision Trees

Training and Visualizing a Decision Tree, Making Predictions, Estimating Class Probabilities, The CART Training Algorithm, Gini Impurity or Entropy, Regularization Hyperparameters, Regression, Instability

8. Ensemble Learning and Random Forests

Voting Classifiers, Bagging and Pasting, Random Patches and Random Subspaces, Random Forests, Boosting, Stacking

9. Dimensionality Reduction

The Curse of Dimensionality, Main Approaches for Dimensionality Reduction, PCA, Kernel PCA, LLE, Other Dimensionality Reduction Techniques

Course Curriculum

Course 5: Deep Learning

1. Introduction to Artificial Neural Networks

From Biological to Artificial Neurons, Implementing MLPs using Keras with TensorFlow Backend, Fine-Tuning Neural Network Hyperparameters

2. Training Deep Neural Networks

The Vanishing / Exploding Gradients Problems, Reusing Pretrained Layers, using Faster Optimizers, Avoiding Overfitting Through Regularization, Practical Guidelines to Train Deep Neural Networks

3. Custom Models and Training with Tensorflow

A Quick Tour of TensorFlow, Using TensorFlow like Numpy, Customizing Models and Training Algorithms, Tensorflow Functions and Graphs

4. Loading and Preprocessing Data with TensorFlow

Introduction to the Data API, TFRecord Format, Preprocessing the Input Features, TF Transform, The TensorFlow Datasets (TDFS) Projects

5. Convolutional Neural Networks

The Architecture of the Visual Cortex, Convolutional Layer, Pooling Layer, CNN Architectures, Classification with Keras, Transfer Learning with Keras, Object Detection, YOLO

6. Recurrent Neural Networks

Recurrent Neurons and Layers, Basic RNNs in TensorFlow, Training RNNs, Deep RNNs, Forecasting a Time Series, LSTM Cell, GRU Cell

Course Curriculum

Course 5: Deep Learning

7. Natural Language Processing

Introduction to Natural Language Processing, Creating a Quiz Using TextBlob, Finding Related Posts with scikit-learn, Generating Shakespearean Text Using Character RNN, Sentiment Analysis, Encoder-Decoder Network for Neural Machine Translation, Attention Mechanisms, Recent Innovations in Language Models

8. Autoencoders and GANs

Efficient Data Representations, Performing PCA with an Under Complete Linear Autoencoder, Stacked Autoencoders, Unsupervised Pre Training Using Stacked Autoencoders, Denoising Autoencoders, Sparse Autoencoders, Variational Autoencoders, Generative Adversarial Networks

9. Reinforcement Learning

Learning to Optimize Rewards, Policy Search, Introduction to OpenAI Gym, Neural Network Policies, Evaluating Actions: The Credit Assignment Problem, Policy Gradients, Markov Decision Processes, Temporal Difference Learning and Q-Learning, Deep Q-Learning Variants, The TF-Agents Library

Projects

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

Projects

- **Analyze Emails**
  Churn the mail activity from various individuals in an open source project development team.
- **Predict bikes rental demand**
  Build a model to predict the bikes demand given the past data.
- **Noise removal from the images**
  Build a model that takes a noisy image as an input and outputs the clean image.
- **Predict which passengers survived in the Titanic shipwreck**
  The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. In this project, you build a model to predict which passengers survived the tragedy.
- **Build a spam classifier**
  Build a model to classify emails as spam or ham using NLTK, and scikit-learn.
- **Build an Image Classifier in Fashion MNIST dataset**
  Classify images from the Fashion MNIST dataset using scikit-learn, and Python.
- **Deploy Machine Learning models to Production using Flask**
  Learn how to deploy a machine learning model as a web application using the Flask framework.

Projects

- **Build an Image Classifier in Fashion MNIST dataset**
  Classify images from the Fashion MNIST dataset using Tensorflow 2, Matplotlib, and Python.
- **Training from Scratch vs Transfer Learning**
  Learn how to train a neural network from scratch to classify data using TensorFlow 2, and how to use the weights of an already trained model to achieve classification to another set of data.
- **Working with Custom Loss Function**
  Create a custom loss function in Keras with TensorFlow 2 backend.
- **Image Classification with Pre-trained Keras models**
  Learn how to access the pre-trained models(here we get pre-trained ResNet model) from Keras of TensorFlow 2 to classify images.
- **Build cats classifier using transfer learning**
  In this project, you will build a basic neural network to classify if a given image is of cat or not using transfer learning technique with Python and Keras.
- **Mask R-CNN with OpenCV for Object Detection**
  Learn how to read a pre-trained TensorFlow model for object detection using OpenCV.
- **Art Generation Project**
  Use TensorFlow 2 to generate an image that is an artistic blend of a content image and style image using Neural Style Transfer.

Projects

- **NYSE Stock Closing Price Prediction using TensorFlow 2 & Keras**
  Predict stock market closing prices for a firm using GRU, a state-of-art deep learning algorithm for sequential data, with Keras and Python.
- **Sentiment Analysis using IMDB dataset**
  Create a sentiment analysis model with the IMDB dataset using TensorFlow 2.
- **Autoencoders for Fashion MNIST**
  Learn how to practically implement the autoencoder, stacking an encoder and decoder using TensorFlow 2, and depict reconstructed output images by the autoencoder model using the Fashion MNIST dataset.
- **Deploy Image Classification Pre-trained Keras model using Flask**
  Learn how to deploy a deep learning model as a web application using the Flask framework.

Course Details and Fees

Please find more information about the course and fees here:
https://cloudxlab.com/course/73/data-science-specialization-eict-iitr

Our Esteemed Customers

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