edureka!
Discover Learning

Data Science Masters Program

Course Curriculum

masters programme

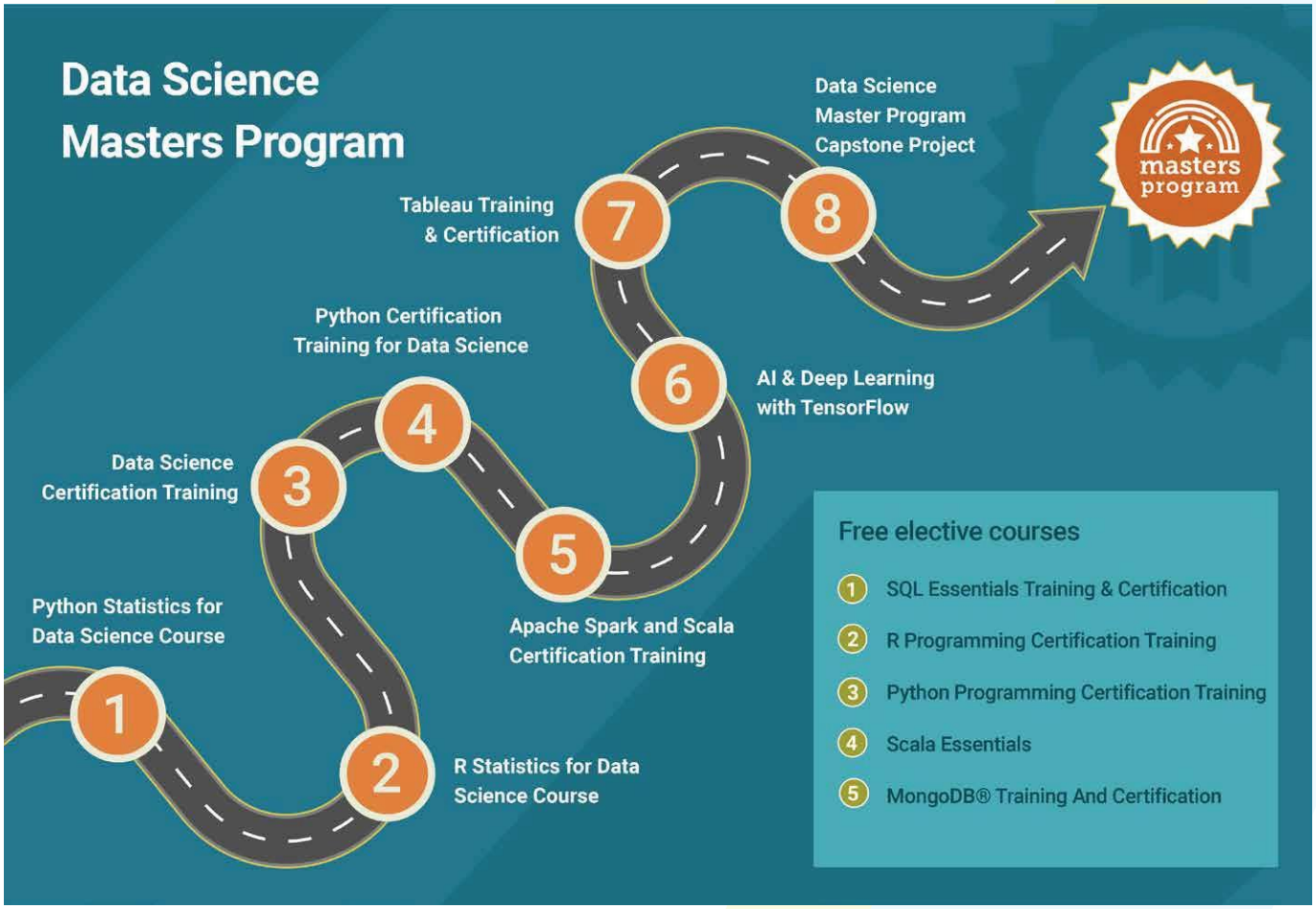
About Edureka

Edureka is a leading e-learning platform providing live instructor-led interactive online training. We cater to professionals and students across the globe in categories like Big Data & Hadoop, Business Analytics, NoSQL Databases, Java & Mobile Technologies, System Engineering, Project Management and Programming.

We have an easy and affordable learning solution that is accessible to millions of learners. With our students spread across countries like the US, India, UK, Canada, Singapore, Australia, Middle East, Brazil and many others, we have built a community of over 1 million learners across the globe.

About The Course

Data Science Masters Program makes you proficient in tools and systems used by Data Science Professionals. It includes training on Statistics, Data Science, Python, Apache Spark & Scala, Tensorflow and Tableau. The curriculum has been determined by extensive research on 5000+ job descriptions across the globe.



Index

Python Statistics for Data Science Course

Course Curriculum

Module 1 : Understanding the Data

Learning Objectives

At the end of this Module, you should be able to:

- Understand various data types
- Learn Various variable types
- List the uses of variable types
- Explain Population and Sample
- Discuss sampling techniques
- Understand Data representation

Topics

- Introduction to Data Types
- Numerical parameters to represent data
- Mean
- Mode
- Median
- Sensitivity
- Information Gain
- Entropy
- Statistical parameters to represent data

Hands-on/Demo

- Estimating mean,median and mode using python
- Calculating Information Gain and Entropy

Module 2 : Probability and its uses

Learning Objectives

At the end of this Module, you should be able to:

- Understand rules of probability
- Learn about dependent and independent events
- Implement conditional, marginal and joint probability using Bayes Theorem
- Discuss probability distribution
- Explain Central Limit Theorem

Topics

- Uses of probability
- Need of probability
- Bayesian Inference
- Density Concepts
- Normal Distribution Curve

Hands-on/Demo

- Calculating probability using python
- Conditional, Joint and Marginal Probability using Python
- Plotting a Normal distribution curve

Module 3 : Statistical Inference

Learning Objectives

At the end of this Module, you should be able to:

- Understand concept of point estimation using confidence margin
- Draw meaningful inferences using margin of error
- Explore hypothesis testing and its different levels

Topics

- Point Estimation
- Confidence Margin
- Hypothesis Testing
- Levels of Hypothesis Testing

Hands-on/Demo

- Calculating and generalizing point estimates using python
- Estimation of Confidence Intervals and Margin of Error

Module 4 : Testing the Data

Learning Objectives

At the end of this module, you should be able to:

- Understand Parametric and Non-parametric Testing
- Learn various types of parametric testing
- Discuss experimental designing
- Explain a/b testing

Topics

- Parametric Test
- Parametric Test Types
- Non- Parametric Test
- Experimental Designing
- A/B testing

Hands-on/Demo

- Perform p test and t tests in python
- A/B testing in python

Module 5 : Data Clustering

Learning Objectives

At the end of this module, you should be able to:

- Understand concept of association and dependence
- Explain causation and correlation
- Learn the concept of covariance
- Discuss Simpson's paradox
- Illustrate Clustering Techniques

Topics

- Association and Dependence
- Causation and Correlation
- Covariance
- Simpson's Paradox
- Clustering Techniques

Hands-on/Demo

- Correlation and Covariance in python
- Hierarchical clustering in python
- K means clustering in python

Module 6 : Regression Modelling

Learning Objectives

At the end of this module, you should be able to:

- Understand the concept of Linear Regression
- Explain Logistic Regression
- Implement WOE
- Differentiate between heteroscedasticity and homoscedasticity
- Learn concept of residual analysis

Topics

- Logistic and Regression Techniques
- Problem of Collinearity
- WOE and IV
- Residual Analysis
- Heteroscedasticity
- Homoscedasticity

Hands-on/Demo

- Perform Linear and Logistic Regression in python
- Analyze the residuals using python

R Statistics for Data Science Course

Course Curriculum

Module 1 : Understanding the Data

Learning Objectives

At the end of this Module, you should be able to: Understand various data types, Learn Various variable types, List the uses of Variable types, Explain Population and Sample, Discuss Sampling techniques and Understand Data representation

Topics

- Introduction to Data Types
- Numerical parameters to represent data
- Mean
- Mode
- Median
- Sensitivity
- Information Gain
- Entropy
- Statistical parameters to represent data

Hands-on/Demo

- Estimating mean, median and mode using R
- Calculating Information Gain and Entropy

Module 2 : Probability and its Uses

Learning Objectives

At the end of this Module, you should be able to: Understand rules of probability, Learn about dependent and independent events, Implement conditional, marginal and joint probability using Bayes Theorem, Discuss probability distribution and Explain Central Limit Theorem

Topics

- Uses of probability
- Need of probability
- Bayesian Inference
- Density Concepts
- Normal Distribution Curve

Hands-on/Demo

- Calculating probability using R
- Conditional, Joint and Marginal Probability using R
- Plotting a Normal distribution curve

Module 3 : Statistical Inference

Learning Objectives

At the end of this Module, you should be able to: Understand the concept of point estimation using confidence margin, Demonstrate the use of Level of Confidence and Confidence Margin, Draw meaningful inferences using margin of error and Explore hypothesis testing and its different levels

Topics

- Point Estimation
- Confidence Margin
- Hypothesis Testing
- Levels of Hypothesis Testing

Hands-on/Demo

- Calculating and generalizing point estimates using R
- Estimation of Confidence Intervals and Margin of Error

Module 4 : Testing the Data

Learning Objectives

At the end of this module, you should be able to: Understand Parametric and Non-Parametric testing, Learn various types of Parametric testing and Explain A/B testing

Topics

- Parametric Test
- Parametric Test Types
- Non- Parametric Test
- A/B testing

Hands-on/Demo

- Perform P test and T tests in R

Module 5 : Data Clustering

Learning Objectives

At the end of this module, you should be able to: Understand the concept of Association and Dependence, Explain Causation and Correlation, Learn the concept of Covariance, Discuss Simpson's paradox and Illustrate Clustering Techniques

Topics

- Association and Dependence
- Causation and Correlation
- Covariance
- Simpson's Paradox
- Clustering Techniques

Hands-on/Demo

- Correlation and Covariance in R
- Hierarchical clustering in R
- K means clustering in R

Module 6 : Regression Modelling

Learning Objectives

At the end of this module, you should be able to: Understand the concept of Linear Regression, Explain Logistic Regression, Implement WOE, Differentiate between heteroscedasticity and homoscedasticity and Learn concept of residual analysis

Topics

- Logistic and Regression Techniques
- Problem of Collinearity
- WOE and IV
- Residual Analysis
- Heteroscedasticity
- Homoscedasticity

Hands-on/Demo

- Perform Linear and Logistic Regression in R
- Analyze the residuals using R
- Calculation of WOE values using R

Data Science Certification Training

Course Curriculum

Module 1 : Introduction to Data Science

Learning Objectives

Get an introduction to Data Science in this module and see how Data Science helps to analyze large and unstructured data with different tools.

Topics

- What is Data Science?
- What does Data Science involve?
- Era of Data Science
- Business Intelligence vs Data Science
- Life cycle of Data Science
- Tools of Data Science
- Introduction to Big Data and Hadoop
- Introduction to R
- Introduction to Spark
- Introduction to Machine Learning

Module 2 : Statistical Inference

Learning Objectives

In this module, you will learn about different statistical techniques and terminologies used in data analysis.

Topics

- What is Statistical Inference?
- Terminologies of Statistics
- Measures of Centers
- Measures of Spread
- Probability
- Normal Distribution
- Binary Distribution

Module 3 : Data Extraction, Wrangling and Exploration

Learning Objectives

Discuss the different sources available to extract data, arrange the data in structured form, analyze the data, and represent the data in a graphical format.

Topics

- Data Analysis Pipeline
- What is Data Extraction
- Types of Data
- Raw and Processed Data
- Data Wrangling
- Exploratory Data Analysis
- Visualization of Data

Hands-on/Demo

- Loading different types of dataset in R
- Arranging the data
- Plotting the graphs

Module 4 : Introduction to Machine Learning

Learning Objectives

Get an introduction to Machine Learning as part of this module. You will discuss the various categories of Machine Learning and implement Supervised Learning Algorithms.

Topics

- What is Machine Learning?
- Machine Learning Use-Cases
- Machine Learning Process Flow
- Machine Learning Categories
- Supervised Learning algorithm: Linear Regression and Logistic Regression

Hands-on/Demo

- Implementing Linear Regression model in R
- Implementing Logistic Regression model in R

Module 5 : Classification Techniques

Learning Objectives

In this module, you should learn the Supervised Learning Techniques and the implementation of various techniques, such as Decision Trees, Random Forest Classifier, etc.

Topics

- What are classification and its use cases?
- What is Decision Tree?
- Algorithm for Decision Tree Induction
- Creating a Perfect Decision Tree
- Confusion Matrix
- What is Random Forest?
- What is Navies Bayes?
- Support Vector Machine: Classification

Hands-on/Demo

- Implementing Decision Tree model in R
- Implementing Linear Random Forest in R
- Implementing Navies Bayes model in R
- Implementing Support Vector Machine in R

Module 6 : Unsupervised Learning

Learning Objectives

Learn about Unsupervised Learning and the various types of clustering that can be used to analyze the data.

Topics

- What is Clustering & its use cases
- What is K-means Clustering?
- What is C-means Clustering?
- What is Canopy Clustering?
- What is Hierarchical Clustering?

Hands-on/Demo

- Implementing K-means Clustering in R
- Implementing C-means Clustering in R
- Implementing Hierarchical Clustering in R

Module 7 : Recommender Engines

Learning Objectives

In this module, you should learn about association rules and different types of Recommender Engines.

Topics

- What is Association Rules & its use cases?
- What is Recommendation Engine & it's working?
- Types of Recommendations
- User-Based Recommendation
- Item-Based Recommendation
- Difference: User-Based and Item-Based Recommendation
- Recommendation use cases

Hands-on/Demo

- Implementing Association Rules in R
- Building a Recommendation Engine in R

Module 8 : Text Mining

Learning Objectives

Discuss Unsupervised Machine Learning Techniques and the implementation of different algorithms, for example, TF-IDF and Cosine Similarity in this Module.

Topics

- The concepts of text-mining
- Use cases
- Text Mining Algorithms
- Quantifying text
- TF-IDF
- Beyond TF-IDF

Hands-on/Demo

- Implementing Bag of Words approach in R
- Implementing Sentiment Analysis on Twitter Data using R

Module 9 : Time Series

Learning Objectives

In this module, you should learn about Time Series data, different component of Time Series data, Time Series modeling - Exponential Smoothing models and ARIMA model for Time Series Forecasting.

Topics

- What is Time Series data?
- Time Series variables
- Different components of Time Series data
- Visualize the data to identify Time Series Components
- Implement ARIMA model for forecasting
- Exponential smoothing models
- Identifying different time series scenario based on which different Exponential Smoothing model can be applied
- Implement respective ETS model for forecasting

Hands-on/Demo

- Visualizing and formatting Time Series data
- Plotting decomposed Time Series data plot
- Applying ARIMA and ETS model for Time Series Forecasting
- Forecasting for given Time period

Module 10 : Deep Learning

Learning Objectives

Get introduced to the concepts of Reinforcement learning and Deep learning in this module. These concepts are explained with the help of Use cases. You will get to discuss Artificial Neural Network, the building blocks for Artificial Neural Networks, and few Artificial Neural Network terminologies.

Topics

- Reinforced Learning
- Reinforcement learning Process Flow
- Reinforced Learning Use cases
- Deep Learning
- Biological Neural Networks
- Understand Artificial Neural Networks
- Building an Artificial Neural Network
- How ANN works
- Important Terminologies of ANN's

Python Certification Training for Data Science

Course Curriculum

Module 1 : Introduction to Python

Learning Objectives

You will get a brief idea of what Python is and touch on the basics

Topics

- Overview of Python
- The Companies using Python
- Different Applications where Python is used
- Discuss Python Scripts on UNIX/Windows
- Values, Types, Variables
- Operands and Expressions
- Conditional Statements
- Loops
- Command Line Arguments
- Writing to the screen

Hands-on/Demo

- Creating "Hello World" code
- Variables
- Demonstrating Conditional Statements
- Demonstrating Loops

Module 2 : Sequences and File Operations

Learning Objectives

Learn different types of sequence structures, related operations and their usage. Also learn diverse ways of opening, reading, and writing to files

Topics

- Python files I/O Functions
- Numbers
- Strings and related operations
- Tuples and related operations
- Lists and related operations
- Dictionaries and related operations
- Sets and related operations

Hands-on/Demo

- Tuple - properties, related operations, compared with a list
- List - properties, related operations
- Dictionary - properties, related operations
- Set - properties, related operations

Module 3 : Deep Dive – Functions, OOPs, Modules, Errors and Exceptions

Learning Objectives

In this Module, you will learn how to create generic python scripts, how to address errors/exceptions in code and finally how to extract/filter content using regex.

Topics

- Functions
- Global Variables
- Variable Scope and Returning Values
- Lambda Functions
- Object-Oriented Concepts
- Standard Libraries
- Modules Used in Python
- Function Parameters
- The Import Statements
- Module Search Path
- Package Installation Ways
- Errors and Exception Handling
- Handling Multiple Exceptions

Hands-on/Demo

- Functions - Syntax, Arguments, Keyword Arguments, Return Values
- Lambda - Features, Syntax, Options, Compared with the Functions
- Sorting - Sequences, Dictionaries, Limitations of Sorting
- Errors and Exceptions - Types of Issues, Remediation
- Packages and Module - Modules, Import Options, sys Path

Module 4 : Introduction to NumPy, Pandas and Matplotlib

Learning Objectives

This Module helps you get familiar with basics of statistics, different types of measures and probability distributions, and the supporting libraries in Python that assist in these operations. Also, you will learn in detail about data visualization.

Topics

- NumPy - arrays
- Operations on arrays
- Indexing slicing and iterating
- Reading and writing arrays on files
- Pandas - data structures & index operations
- Reading and Writing data from Excel/CSV formats into Pandas

- matplotlib library
- Grids, axes, plots
- Markers, colours, fonts and styling
- Types of plots - bar graphs, pie charts, histograms
- Contour plots

Hands-on/Demo

- NumPy library- Creating NumPy array, operations performed on NumPy array
- Pandas library- Creating series and dataframes, Importing and exporting data
- Matplotlib - Using Scatterplot, histogram, bar graph, pie chart to show information, Styling of Plot

Module 5 : Data Manipulation

Learning Objectives

Through this Module, you will understand in detail about Data Manipulation

Topics

- Basic Functionalities of a data object
- Merging of Data objects
- Concatenation of data objects
- Types of Joins on data objects
- Exploring a Dataset
- Analysing a dataset

Hands-on/Demo

- Pandas Function- Ndim(), axes(), values(), head(), tail(), sum(), std(), iteritems(), iterrows(), itertuples()
- GroupBy operations
- Aggregation
- Concatenation
- Merging
- Joining

Module 6 : Introduction to Machine Learning with Python

Learning Objectives

In this module, you will learn the concept of Machine Learning and its types.

Topics

- Python Revision (numpy, Pandas, scikit learn, matplotlib)
- What is Machine Learning?
- Machine Learning Use-Cases
- Machine Learning Process Flow
- Machine Learning Categories
- Linear regression
- Gradient descent

Hands-on/Demo

- Linear Regression – Boston Dataset

Module 7 : Supervised Learning - I

Learning Objectives

In this module, you will learn Supervised Learning Techniques and their implementation, for example, Decision Trees, Random Forest Classifier etc.

Topics

- What are Classification and its use cases?
- What is Decision Tree?
- Algorithm for Decision Tree Induction
- Creating a Perfect Decision Tree
- Confusion Matrix
- What is Random Forest?

Hands-on/Demo

- Implementation of Logistic regression
- Decision tree
- Random forest

Module 8 : Dimensionality Reduction

Learning Objectives

In this module, you will learn about the impact of dimensions within data. You will be taught to perform factor analysis using PCA and compress dimensions. Also, you will be developing LDA model.

Topics

- Introduction to Dimensionality
- Why Dimensionality Reduction
- PCA
- Factor Analysis
- Scaling dimensional model
- LDA

Hands-on/Demo

- PCA
- Scaling

Module 9 : Supervised Learning - II

Learning Objectives

In this module, you will learn Supervised Learning Techniques and their implementation, for example, Decision Trees, Random Forest Classifier etc.

Topics

- What is Naïve Bayes?
- How Naïve Bayes works?
- Implementing Naïve Bayes Classifier
- What is Support Vector Machine?
- Illustrate how Support Vector Machine works?
- Hyperparameter Optimization
- Grid Search vs Random Search
- Implementation of Support Vector Machine for Classification

Hands-on/Demo

- Implementation of Naïve Bayes, SVM

Module 10 : Unsupervised Learning

Learning Objectives

In this module, you will learn about Unsupervised Learning and the various types of clustering that can be used to analyze the data.

Topics

- What is Clustering & its Use Cases?
- What is K-means Clustering?
- How does K-means algorithm work?
- How to do optimal clustering
- What is C-means Clustering?
- What is Hierarchical Clustering?
- How Hierarchical Clustering works?

Hands-on/Demo

- Implementing K-means Clustering
- Implementing Hierarchical Clustering

Module 11 : Association Rules Mining and Recommendation Systems

Learning Objectives

In this module, you will learn Association rules and their extension towards recommendation engines with Apriori algorithm.

Topics

- What are Association Rules?
- Association Rule Parameters
- Calculating Association Rule Parameters
- Recommendation Engines
- How does Recommendation Engines work?
- Collaborative Filtering
- Content-Based Filtering

Hands-on/Demo

- Apriori Algorithm
- Market Basket Analysis

Module 12 : Reinforcement Learning

Learning Objectives

In this module, you will learn about developing a smart learning algorithm such that the learning becomes more and more accurate as time passes by. You will be able to define an optimal solution for an agent based on agent-environment interaction.

Topics

- What is Reinforcement Learning
- Why Reinforcement Learning
- Elements of Reinforcement Learning
- Exploration vs Exploitation dilemma
- Epsilon Greedy Algorithm
- Markov Decision Process (MDP)
- Q values and V values
- Q – Learning
- α values

Hands-on/Demo

- Calculating Reward
- Discounted Reward
- Calculating Optimal quantities
- Implementing Q Learning
- Setting up an Optimal Action

Module 13 : Time Series Analysis

Learning Objectives

In this module, you will learn about Time Series Analysis to forecast dependent variables based on time. You will be taught different models for time series modeling such that you analyze a real time-dependent data for forecasting.

Topics

- What is Time Series Analysis?
- Importance of TSA
- Components of TSA
- White Noise
- AR model
- MA model
- ARMA model
- ARIMA model
- Stationarity
- ACF & PACF

Hands-on/Demo

- Checking Stationarity
- Converting a non-stationary data to stationary
- Implementing Dickey-Fuller Test
- Plot ACF and PACF
- Generating the ARIMA plot
- TSA Forecasting

Module 14 : Model Selection and Boosting

Learning Objectives

In this module, you will learn about selecting one model over another. Also, you will learn about Boosting and its importance in Machine Learning. You will learn on how to convert weaker algorithms into stronger ones.

Topics

- What is Model Selection?
- The need for Model Selection
- Cross-Validation
- What is Boosting?
- How Boosting Algorithms work?
- Types of Boosting Algorithms
- Adaptive Boosting

Hands-on/Demo

- Cross-Validation
- AdaBoost

Apache Spark and Scala Certification Training

Course Curriculum

About The Course

Edureka's Apache Spark and Scala Certification Training is designed to provide you the knowledge and skills that are required to become a successful Spark Developer and prepare you for the Cloudera Hadoop and Spark Developer Certification Exam (CCA175). Throughout the Apache Spark Training, you will get an in-depth knowledge on Apache Spark and the Spark Ecosystem, which includes Spark RDD, Spark SQL, Spark MLlib and Spark Streaming.You will also get comprehensive knowledge on Scala Programming language, HDFS, Sqoop, FLume, Spark GraphX and Messaging System such as Kafka.

Module 1 : Introduction to Big Data Hadoop and Spark

Learning Objectives

Understand Big Data and its components such as HDFS. You will learn about the Hadoop Cluster Architecture and you will also get an introduction to Spark and you will get to know about the difference between batch processing and real-time processing.

Topics

- What is Big Data?
- Big Data Customer Scenarios
- Limitations and Solutions of Existing Data Analytics Architecture with Uber Use Case
- How Hadoop Solves the Big Data Problem?
- What is Hadoop?
- Hadoop's Key Characteristics
- Hadoop Ecosystem and HDFS
- Hadoop Core Components
- Rack Awareness and Block Replication
- YARN and its Advantage
- Hadoop Cluster and its Architecture
- Hadoop: Different Cluster Modes
- Big Data Analytics with Batch & Real-time Processing
- Why Spark is needed?
- What is Spark?
- How Spark differs from other frameworks?
- Spark at Yahoo!

Module 2 : Introduction to Scala and Apache Spark

Learning Objectives

Learn the basics of Scala that are required for programming Spark applications. You will also learn about the basic constructs of Scala such as variable types, control structures, collections such as Array, ArrayBuffer, Map, Lists, and many more.

Topics

- What is Scala?
- Why Scala for Spark?
- Scala in other Frameworks
- Introduction to Scala REPL
- Basic Scala Operations
- Variable Types in Scala
- Control Structures in Scala
- Foreach loop, Functions and Procedures
- Collections in Scala- Array
- ArrayBuffer, Map, Tuples, Lists, and more

Hands On

- Scala REPL Detailed Demo

Module 3 : Functional Programming and OOPs Concepts in Scala

Learning Objectives

In this module, you will learn about object-oriented programming and functional programming techniques in Scala.

Topics

- Functional Programming
- Anonymous Functions
- Getters and Setters
- Properties with only Getters
- Singletons
- Overriding Methods
- Higher Order Functions
- Class in Scala
- Custom Getters and Setters
- Auxiliary Constructor and Primary Constructor
- Extending a Class
- Traits as Interfaces and Layered Traits

Hands On

- OOPs Concepts
- Functional Programming

Module 4 : Deep Dive into Apache Spark Framework

Learning Objectives

Understand Apache Spark and learn how to develop Spark applications. At the end, you will learn how to perform data ingestion using Sqoop.

Topics

- Spark's Place in Hadoop Ecosystem
- Spark Components & its Architecture
- Spark Deployment Modes
- Introduction to Spark Shell
- Writing your first Spark Job Using SBT

- Submitting Spark Job
- Spark Web UI
- Data Ingestion using Sqoop

Hands On

- Building and Running Spark Application
- Spark Application Web UI
- Configuring Spark Properties
- Data ingestion using Sqoop

Module 5 : Playing with Spark RDDs

Learning Objectives

Get an insight of Spark - RDDs and other RDD related manipulations for implementing business logics (Transformations, Actions and Functions performed on RDD).

Topics

- Challenges in Existing Computing Methods
- Probable Solution & How RDD Solves the Problem
- What is RDD, It's Functions, Transformations & Actions?
- Data Loading and Saving Through RDDs
- Key-Value Pair RDDs
- Other Pair RDDs o RDD Lineage
- RDD Lineage
- RDD Persistence
- WordCount Program Using RDD Concepts
- RDD Partitioning & How It Helps Achieve Parallelization
- Passing Functions to Spark

Hands On/Demo

- Loading data in RDDs
- RDD Transformations
- RDD Partitions
- Saving data through RDDs
- RDD Actions and Functions
- WordCount through RDDs

Module 6 : DataFrames and Spark SQL

Learning Objectives

In this module, you will learn about SparkSQL which is used to process structured data with SQL queries, data-frames and datasets in Spark SQL along with different kind of SQL operations performed on the data-frames. You will also learn about the Spark and Hive integration.

Topics

- Need for Spark SQL
- What is Spark SQL?
- Spark SQL Architecture
- SQL Context in Spark SQL
- User Defined Functions
- Data Frames & Datasets
- Interoperating with RDDs
- JSON and Parquet File Formats
- Loading Data through Different Sources
- Spark – Hive Integration

Hands On/Demo

- Spark SQL – Creating Data Frames
- Loading and Transforming Data through Different Sources
- Stock Market Analysis
- Spark-Hive Integration

Module 7 : Machine Learning using Spark MLlib

Learning Objectives

Learn why machine learning is needed, different Machine Learning techniques/algorithms, and SparK MLlib.

Topics

- Why Machine Learning?
- What is Machine Learning?

- Where Machine Learning is Used?
- Face Detection: USE CASE
- Different Types of Machine Learning Techniques
- Introduction to MLlib
- Features of MLlib and MLlib Tools
- Various ML algorithms supported by MLlib

Module 8 : Deep Dive into Spark MLlib

Learning Objectives

Implement various algorithms supported by MLlib such as Linear Regression, Decision Tree, Random Forest and many more.

Topics

- Supervised Learning - Linear Regression, Logistic Regression, DecisionmTree, Random Forest
- Unsupervised Learning - K-Means Clustering & How It Works with MLlib
- Analysis on US Election Data using MLlib (K-Means)

Hands-On

- Machine Learning MLlib
- Linear Regression
- Decision Tree
- K- Means Clustering
- Logistic Regression
- Random Forest

Module 9 : Understanding Apache Kafka & Apache Flume

Learning Objectives

Understand Kafka and its Architecture. Also, learn about Kafka Cluster, how to configure different types of Kafka Cluster. Get introduced to Apache Flume, its architecture and how it is integrated with Apache Kafka for event processing. At the end, learn how to ingest reaming data using flume.

Topics

- Need for Kafka
- What is Kafka?
- Core Concepts of Kafka
- Kafka Architecture
- Where is Kafka Used?
- Understanding the Components of Kafka Cluster
- Configuring Kafka Cluster
- Need of Apache Flume

- What is Apache Flume?
- Basic Flume Architecture
- Flume Sources
- Flume Sinks
- Flume Channels
- Flume Configuration
- Integrating Apache Flume and Apache Kafka

Hands-On

- Configuring Single Node Single Broker Cluster
- Configuring Single Node Multi Broker Cluster
- Producing and consuming messages
- Flume Commands
- Setting up Flume Agent
- Streaming Twitter Data into HDFS

Module 10 : Apache Spark Streaming- Processing Multiple Batches

Learning Objectives

Work on Spark streaming which is used to build scalable fault-tolerant streaming applications.

Also, learn about DStreams and various Transformations performed on the streaming data.

You will get to know about commonly used streaming operators such as Sliding

Window Operators and Stateful Operators.

Topics

- Drawbacks in Existing Computing Methods
- Why Streaming is Necessary?
- What is Spark Streaming?
- Spark Streaming Features
- Spark Streaming Workflow
- How Uber Uses Streaming Data
- Streaming Context & DStreams
- Transformations on DStreams

- Describe Windowed Operators and Why it is Useful
- Important Windowed Operators
- Slice, Window and ReduceByWindow Operators
- Stateful Operators

Module 11 : Apache Spark Streaming- Data Sources

Learning Objectives

In this module, you will learn about the different streaming data sources such as Kafka and flume. At the end of the module, you will be able to create a spark streaming application.

Topics

- Apache Spark Streaming: Data Sources
- Streaming Data Source Overview
- Apache Flume and Apache Kafka Data Sources
- Example: Using a Kafka Direct Data Source
- Perform Twitter Sentimental Analysis Using Spark Streaming

Hands-On

- Different Streaming Data Sources

Module 12 : In Class Project

Learning Objectives

Work on an end-to-end Financial domain project covering all the major concepts of Spark taught during the course.

Module 13 : Spark GraphX (Self-Paced)

Learning Objectives

In this module, you will be learning the key concepts of Spark GraphX programming and operations along with different GraphX algorithms and their implementations.

AI & Deep Learning with TensorFlow

Course Curriculum

Module 1 : Introduction to Deep Learning

Learning Objectives

In this module, you'll get an introduction to Deep Learning and understand how Deep Learning solves problems which Machine Learning cannot. Understand fundamentals of Machine Learning and relevant topics of Linear Algebra and Statistics.

Topics

- Deep Learning: A revolution in Artificial Intelligence
- Limitations of Machine Learning
- What is Deep Learning?
- Advantage of Deep Learning over Machine learning
- 3 Reasons to go for Deep Learning
- Real-Life use cases of Deep Learning
- Review of Machine Learning: Regression, Classification, Clustering, Reinforcement Learning, Underfitting and Overfitting, Optimization

Hands-On

- Implementing a Linear Regression model for predicting house prices from Boston dataset
- Implementing a Logistic Regression model for classifying Customers based on a Automobile purchase dataset

Module 2 : Understanding Neural Networks with TensorFlow

Learning Objectives

In this module, you'll get an introduction to Neural Networks and understand it's working i.e. how it is trained, what are the various parameters considered for its training and the activation functions that are applied.

Topics

- How Deep Learning Works?
- Activation Functions
- Illustrate Perceptron
- Training a Perceptron
- Important Parameters of Perceptron
- What is TensorFlow?
- TensorFlow code-basics
- Graph Visualization
- Constants, Placeholders, Variables
- Creating a Model
- Step by Step - Use-Case Implementation

Hands-On

- Building a single perceptron for classification on SONAR dataset

Module 3 : Deep dive into Neural Networks with TensorFlow

Learning Objectives

In this module, you'll understand backpropagation algorithm which is used for training Deep Networks. You will know how Deep Learning uses neural network and backpropagation to solve the problems that Machine Learning cannot.

Topics

- Understand limitations of a Single Perceptron
- Understand Neural Networks in Detail
- Illustrate Multi-Layer Perceptron

- Backpropagation – Learning Algorithm
- Understand Backpropagation – Using Neural Network Example
- MLP Digit-Classifier using TensorFlow
- TensorBoard

Hands-On

- Building a multi-layered perceptron for classification of Hand-written digits

Module 4 : Master Deep Networks

Learning Objectives

In this module, you'll get started with the TensorFlow framework. You will understand how it works, its various data types & functionalities. In addition, you will create an image classification model.

Topics

- Why Deep Networks
- Why Deep Networks give better accuracy?
- Use-Case Implementation on SONAR dataset
- Understand How Deep Network Works?
- How Backpropagation Works?
- Illustrate Forward pass, Backward pass
- Different variants of Gradient Descent
- Types of Deep Networks

Hands-On

- Building a multi-layered perceptron for classification on SONAR dataset

Module 5 : Convolutional Neural Networks (CNN)

Learning Objectives

In this module, you'll understand convolutional neural networks and its applications. You will learn the working of CNN, and create a CNN model to solve a problem.

Topics

- Introduction to CNNs
- CNNs Application
- Architecture of a CNN
- Convolution and Pooling layers in a CNN
- Understanding and Visualizing a CNN

Hands-On

- Building a convolutional neural network for image classification. The model should predict the difference between 10 categories of images.

Module 6: Recurrent Neural Networks (RNN)

Learning Objectives

In this module, you'll understand Recurrent Neural Networks and its applications. You will understand the working of RNN, how LSTM are used in RNN, what is Recursive Neural Tensor Network Theory, and finally you will learn to create a RNN model.

Topics

- Introduction to RNN Model
- Application use cases of RNN
- Modelling sequences
- Training RNNs with Backpropagation
- Long Short-Term memory (LSTM)
- Recursive Neural Tensor Network Theory
- Recurrent Neural Network Model

Hands-On

- Building a recurrent neural network for SPAM prediction.

Module 7 : Restricted Boltzmann Machine (RBM) and Autoencoders

Learning Objectives

In this module, you'll understand RBM & Autoencoders along with their applications. You will understand the working of RBM & Autoencoders, illustrate Collaborative Filtering using RBM and understand what are Deep Belief Networks.

Topics

- Restricted Boltzmann Machine
- Applications of RBM
- Collaborative Filtering with RBM
- Introduction to Autoencoders
- Autoencoders applications
- Understanding Autoencoders

Hands-On

- Building a Autoencoder model for classification of handwritten images extracted from the MNIST Dataset

Module 8 : Keras API

Learning Objectives

In this module, you'll understand how to use Keras API for implementing Neural Networks. The goal is to understand various functions and features that Keras provides to make the task of neural network implementation easy.

Topics

- Define Keras
- How to compose Models in Keras
- Sequential Composition
- Functional Composition
- Predefined Neural Network Layers
- What is Batch Normalization
- Saving and Loading a model with Keras
- Customizing the Training Process

Topics

- Using TensorBoard with Keras
- Use-Case Implementation with Keras

Hands-On

- Build a model using Keras to do sentiment analysis on twitter data reactions on GOP debate in Ohio

Module 9 : TFLearn API

Learning Objectives

In this module, you'll understand how to use TFLearn API for implementing Neural Networks. The goal is to understand various functions and features that TFLearn provides to make the task of neural network implementation easy.

Topics

- Define TFLearn
- Composing Models in TFLearn
- Sequential Composition
- Functional Composition
- Predefined Neural Network Layers
- What is Batch Normalization
- Saving and Loading a model with TFLearn
- Customizing the Training Process
- Using TensorBoard with TFLearn
- Use-Case Implementation with TFLearn

Hands-On

- Build a recurrent neural network using TFLearn to do image classification on hand-written digits

Module 10 : In-Class Project

Learning Objectives

In this module, you should learn how to approach and implement a project end to end. The instructor will share his industry experience and related insights that will help you kickstart your career in this domain. In addition, we will be having a QA and doubt clearing session for you.

Topics

- How to approach a project?
- Hands-On project implementation
- What Industry expects?
- Industry insights for the Machine Learning domain
- QA and Doubt Clearing Session

Tableau Training & Certification

Course Curriculum

Module 1 : I Introduction to Data Visualization

Learning Objectives

Identify the prerequisites, goal, objectives, methodology, material, and agenda for the course. Discuss the basic of Data Visualization. Get a brief idea about Tableau, establish connection with the dataset, perform Joins operation on the data set.

Topics

- Data Visualization
- Introducing Tableau 10.0
- Establishing Connection
- Joins and Union
- Data Blending

Hands-On

- Establishing connection with the files, Introducing important UI components (ShowMe, Fit Axes)
- Perform Cross Joins between the dataset

Module 2 : Visual Analytics

Learning Objectives

Manage extracts and metadata (by creating hierarchy and folders). Describe what is Visual Analytics, why to use it, and it's various scopes. Explain aggregating and disaggregating data and how to implement data granularity using marks card on aggregated data. Describe what is highlighting, with the help of a use-case. Illustrate basic graphs including bar graph, line graph, pie chart, dual axis graph, and area graph with dual axis.

Topics

- Managing Extracts
- Managing Metadata
- Visual Analytics
- Data Granularity using Marks Card
- Highlighting
- Introduction to basic graphs

Hands-On

- Creating Extracts, Hierarchy, Folders
- All the features of Marks Card Shelve with use case provided
- Power of Highlighting in the visualization using the Use-case
- How to create basic graphs in Tableau10.x

Module 3 : Visual Analytics in depth I

Learning Objectives

Perform sorting techniques including quicksort, using measures, using header and legend, and sorting using pill with the help of a use case. Master yourself into various filtering techniques such as Parametrized filtering, Quick Filter, Context Filter. Learn about various filtering option available with the help of use case and different scenarios. Illustrate grouping using data-window, visual grouping, and Calculated Grouping (Static and Dynamic). Illustrate some more graphical visualization including Heat Map, Circle Plot, Scatter Plot, and Tree Maps.

- Sorting.
- Filtering.
- Grouping
- Graphical Visualization

Hands-On

- Quick Sort, Sorting using measure, Sorting using header and legends, sorting using pill(use-case).
- Filtering Use cases covering different options (General, Wildcard, Conditional).
- Interactive Filter, Quick Filter, Context Filter.
- Grouping using Data Window, Visual Grouping, Calculated Grouping (Static and Dynamic).

Module 4 : Visual Analytics in depth II

Learning Objectives

Explain the basic concepts of sets followed by Creating sets using Marks Card, computation sets and combined sets. Describe the concepts of forecasting with the help of Forecasting problem as a use-case. Discuss the basic concept of clustering in Tableau. Add Trend lines and reference line to your visualization.Discuss about Parameter in depth using Sets and Filter

Topics

- Sets
- Forecasting
- Clustering
- Trend Lines.
- Reference Lines.
- Parameters

Hands-On

- Create sets using marks card, Computation sets, and Combined sets
- Forecasting using Precise Range
- Methods of clustering
- Adding trend line and reference line (along with various options available for them)
- Parameter using sets and filter

Module 5 : Dashboard and Stories

Learning Objectives

Describe the basic concepts of Dashboard and its UI. Build a dashboard by adding sheets and object into it. Modify the view and layout. Edit your dashboard, how it should appear on phones or tablets. Create an interactive dashboard using actions (filter, highlighting, URL). Create stories for your Visualization and Dashboards.

Topics

- Introduction to Dashboard.
- Creating a Dashboard Layout.
- Designing Dashboard for Devices.
- Dashboard Interaction - Using Action.
- Introduction to Story Point.

Hands-On

- Creating Dashboard and learning its UI component.
- Changing the layout of the dashboard.
- Using Device Designer to create dashboard for devices.
- Create an interactive dashboard using actions (Filter, Highlight, URL).
- Creating story with dashboard.

Module 6 : Mapping

Learning Objectives

Map the coordinates on the map, plot geographic data, and use layered view to get the street view of the area. Edit the ambiguous and unrecognized location plotted on the map. Customize territory in a polygon map. Connect to the WMS Server, use a WMS background map and saving it. Add a background image and generate its coordinate and plot the points.

Topics

- Introduction to Maps.
- Editing Unrecognized Locations.
- Custom Geocoding.
- Polygon Maps.
- Web Mapping Services.
- Background Images.

Hands-On

- Plot the coordinate points on the map, plotting the geographic data, Street View using the layered view.
- Editing Unrecognized and ambiguous location
- Custom Geocoding.
- Creating a custom territory, building a polygon map.

- Establishing connection with the WMS Server a WMS background map and saving it.
- Adding a background image and generate coordinates and finally plotting points.

Module 7 : Calculation

Learning Objectives

Perform Calculations using various types of functions such as Number, String, Date, Logical, and Aggregrate. In addition, you will get to know about Quick Table Calculation. Cover the following LOD expressions – Fixed, Included, and Excluded.

Topics

- Introduction to Calculation : Number Functions, String Functions , Date Functions, Logical Functions, Aggregate Functions.
- Introduction to Table Calculation.
- Introduction to LOD expression : Fixed LOD , Included LOD, Excluded LOD

Hands-On

- All Functions (Number, String, Date, Logical, Aggregate)
- Table Calculation.
- LOD expressions.

Module 8 : LOD Problem Sets & Hands on

Learning Objectives

Tackle complex scenarios by using LOD expressions.

Hands-On

- Use Case I - Count Customer by Order.
- Use Case II - Profit per Business Day.
- Use Case III - Comparative Sales.
- Use Case IV - Profit Vs Target
- Use Case V - Finding the second order date.
- Use Case VI - Cohort Analysis

Module 9 : Charts

Topics

- Box and Whisker's Plots
- Gantt Charts
- Waterfall Charts
- Pareto Charts
- Control Charts
- Funnel Charts

Hands-On

- Extensive hands-on on the above topics

Module 10 : Integrating Tableau with R and Hadoop

Learning Objectives

You will know the basics of Big Data, Hadoop, and R. You will discuss the integration between Hadoop and R and will integrate R with Tableau. In addition, you will get to publish your workbook on Tableau Server.

Topics

- Introduction to Big Data
- Introduction to Hadoop
- Introduction to R
- Integration among R and Hadoop
- Calculating measure using R
- Integrating Tableau with R
- Integrated Visualization using Tableau

Data Science Master Program
Capstone Project

Course Curriculum

Auto Insurance Case Study

Learning Objectives

The capstone project will provide you with a business case. You will need to solve this by applying all the skills you've learned in the courses of the master's program. This Capstone project will require you to apply the following skills

Data Exploration

- Checking Data Size
- Note the important features

Data Wrangling

- Handling Imbalanced Data
- MetaData Creation
- Statistics on the Data
- Identify Missing Variable
- Rectify Missing Variable
- One Hot Encoding
- Scaling: Standard Scaler & Min Max Scaler

Data Exploration

- Data Visualization

Machine Learning

- PCA
- Logistic Regression
- Generating F1 Score Metric
- Linear SVC Classifier
- XG Boost Classifier
- AdaBoost Classifier

Deep Learning Learning

- MLP Classifier
- MLP Classifier with Cross Validation