

Big Data Architect Masters Program

Course Curriculum



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

Edureka's Masters Program offers an in-depth knowledge of Hadoop ecosystem, Deep Learning using Tensor Flow and real-time processing using Spark and NoSQL database technology. As part of this program you get access to 11 specialization courses and 10+ projects. The entire program is a structured learning path recommended by leading industry experts.

Big Data Architect Masters Program

- 1 Python or JAVA
- 2 Fundamentals of Statistics
- 3 Big Data Hadoop Certification
- 4 Apache Spark Certification
- 5 AI & Deep Learning with TensorFlow



- 1 Cassandra Certification Training
- 2 MongoDB Certification Training
- 3 Talend Certification Training
- 4 Fundamentals of Scala
- 5 Tableau Certification & Training

Index

| | | |
|----|----------------------------------------|----|
| 1. | Python Essentials | 03 |
| 2. | Java Essentials | 07 |
| 3. | Fundamentals of Statistics | 10 |
| 4. | Big Data Hadoop Certification Training | 12 |
| 5. | Apache Spark Certification Training | 23 |
| 6. | AI & Deep Learning with TensorFlow | 29 |

Python Essentials

Course Curriculum

About The Course

Python has been one of the premier, flexible, and powerful open-source language that is easy to learn, easy to use, and has powerful libraries for data manipulation and analysis. For over a decade, Python has been used in scientific computing and highly quantitative domains such as finance, oil and gas, physics, and signal processing. As of today, it is the most preferred language for Artificial Intelligence, Robotics, Web Development and DevOps.

Module 1 : Introduction to Python

Learning Objectives

At the end of this Module, you should be able to understand Python – an Object oriented Programming Language, List the Users of Python for Data Analytics, Define Identifiers and Indentation, List Operations on Strings and Numbers, Run a Python Script.

Topics

- ✓ Get an overview of Python
- ✓ Learn about Interpreted Languages
- ✓ List the Advantages/Disadvantages of Python
- ✓ Explore Pydoc
- ✓ Explore Python Editors and IDEs
- ✓ Start Python
- ✓ Discuss Interpreter PATH
- ✓ Run a Python Script
- ✓ Discuss Python Scripts on UNIX/Windows

Module 2 : Sequences and File Operations

Learning Objectives

At the end of this module, you will be able to Define Reserved Keywords and Command Line Arguments, Describe how to Get User Input from Keyboard, Describe Flow Control and Sequences, Practice Working with Files, Define and Describe Dictionaries and Sets.

Topics

- ✓ Lists
- ✓ Iterating through a sequence
- ✓ Tuples
- ✓ Functions for all sequences
- ✓ Indexing and Slicing
- ✓ Using enumerate()
- ✓ Generator expressions
- ✓ Operators and keywords for sequences
- ✓ Dictionaries and sets
- ✓ The xrange()function
- ✓ Working with files
- ✓ List comprehensions
- ✓ Modes of opening a file
- ✓ File methods
- ✓ File attributes

Module 3 : Deep Dive – Functions, Sorting, Errors and Exception, Regular Expressions and Package

Learning Objectives

At the end of this Module, you should be able to explain Functions and various forms of Function Arguments, explain Standard Library, define modules, describe Zip Archives and Packaging.

Topics

- ✓ Functions
- ✓ Function Parameters
- ✓ Global variables
- ✓ Variable scope and Returning Values
- ✓ Errors and Exception Handling
- ✓ Sorting dictionaries
- ✓ Sorting lists in place
- ✓ Sorting
- ✓ Alternate Keys
- ✓ Lambda Functions
- ✓ Sorting collections of collections
- ✓ Handling multiple exceptions
- ✓ The standard exception hierarchy
- ✓ Using Modules

Module 4: Object Oriented Programming in Python

Learning Objectives

At the end of this Module, you should be able to implement Regular Expression and its Basic Functions; Use Classes, Objects, and Attributes, Develop applications based on Object Oriented Programming and Methods

Topics

- ✓ The sys Module
- ✓ Interpreter information
- ✓ STDIO
- ✓ Launching external programs
- ✓ Paths
- ✓ Directories and filenames
- ✓ Walking directory trees
- ✓ Instance methods
- ✓ Class methods and data
- ✓ Private methods and Inheritance
- ✓ Math Function
- ✓ Random Numbers
- ✓ Dates and Times
- ✓ Zipped Archives
- ✓ Introduction to Python Classes
- ✓ Defining Classes
- ✓ Initializes
- ✓ Properties
- ✓ Static methods

Module 5 : Debugging, Databases, and Project Skeletons

Learning Objectives

At the end of this Module, you should be able to debug python scripts using pdb, debug python scripts using IDE, classify Errors, develop Unit Tests, create project Skeletons, implement Database using SQLite and perform CRUD operations on SQLite database.

Topics

- ✓ Debugging
- ✓ Dealing with errors
- ✓ Using unit tests
- ✓ Project Skeleton
- ✓ Required packages
- ✓ Creating the Skeleton
- ✓ Using the skeleton
- ✓ Creating a database with SQLite 3
- ✓ CRUD operations
- ✓ Creating a database object.
- ✓ Project Directory
- ✓ Final Directory Structure
- ✓ Testing your set up

JAVA Essentials

Course Curriculum

About The Course

The 'Fundamentals of Java' course is designed to train the participants in the Java language to become a Java programmer. This course will cover the basic concepts of Java like Data Handling and Functions, Object Oriented Programming in Java, Packages and Multi threading, and Collections.

Module 1 : Introduction to Java

Learning Objectives

At the end of this module, you will be able to understand the advantages of Java, where Java is used, understand how memory management is handled in Java, create a Java project in Eclipse and execute, implement if..else construct in Java, develop codes using various data types in Java, implement various loops.

Topics

- ✓ Introduction to JavaT
- ✓ Bytecode
- ✓ Class Files
- ✓ Compilation Process
- ✓ If conditions
- ✓ Loops - for, while and do while
- ✓ Data types and Operations

Module 2 : Data Handling and Functions

Learning Objectives

At the end of this module, you will be able to implement Single and Multi-dimensional array, declare and Define Function, call Functions by value and by reference, implement Method Overloading and Use String data-type and String-buffer.

Topics

- ✓ Functions
- ✓ Function with Arguments
- ✓ Function Overloading
- ✓ Concept of Static Polymorphism
- ✓ String Handling -String
- ✓ String buffer Classes
- ✓ Arrays - Single Dimensional and Multidimensional arrays

Module 3 : Object Oriented Programming in Java

Learning Objectives

At the end of this module, you will be able to implement classes and objects in Java, create class constructors, overload constructors, inherit classes and create sub-classes, implement abstract classes and methods, and use static keyword.

Topics

- ✓ Concept of Object Orientation
- ✓ Constructors with Arguments
- ✓ Attributes and Methods
- ✓ Inheritance
- ✓ Classes and Objects
- ✓ Abstract
- ✓ Default Constructors
- ✓ Final and Static

Module 4 : Data Handling and Functions

Learning Objectives

At the end of this module, you will be able to implement interface and use it, extend interface with other interface, create package and name it Import packages while creating a new class, understand various exceptions, handle exception using try catch block, handle exception using throw and throws keyword, implement threads using thread class and runnable interface and Understand and implement multi threading.

Topics

- ✓ Packages and Interfaces
- ✓ Access Specifiers
- ✓ Package
- ✓ Exception Handling
- ✓ Multi-Threading

Module 5 : Collections

Learning Objectives

At the end of this module, you will be able to identify, and use important Inbuilt Java Packages like java.lang, java.io, java.util etc, use Wrapper classes, understand collections framework, implement logic using ArrayList and Vector and Queue, use set, HashSet and TreeSet and implement logic using Map HashMap and Hashtable.

Topics

- ✓ Wrapper Classes and Inner Classes
- ✓ Applet Programs
- ✓ Collections: ArrayList, Vector, HashSet

Fundamentals of Statistics

Course Curriculum

About The Course

The self-paced Fundamentals of Statistics Course is designed for the learners to understand and implement various statistical techniques. These techniques are explained using dedicated examples. The use case is taken up at the end of each module and insights are gathered, thus at the end of the course we have a Project which is consistently worked upon throughout the course.

Module 1 : Introduction to Statistics and Basic Probability

Learning Objectives

At the end of this module, you will be able to understand Skewness, Modality, Measures of Center, Measures of Spread etc. You will also understand the relationship between these terminologies. You will also be able to analyze airlines data set to gather insights.

Topics

- ✓ Statistics & Basic Probability
- ✓ Sampling Methods
- ✓ Measures of Center
- ✓ Measures of Spread

Module 2 : Basic Probability, Conditional Probability and Bayesian Inference

Learning Objectives

At the end of this module, you will be able to understand the rules of probability, learn about Disjoint and Independent events, understand the concept of probability, and implement these concepts on a case-study. You will also learn and implement Bayes' Theorem and implement Bayes' theorem on a case-study.

Topics

- ✓ Conditional Probability & Bayesian Inference
- ✓ Definitions
- ✓ Examples
- ✓ Terms
- ✓ Concepts & Applications

Module 3 : Distributions and Regression Modeling

Learning Objectives

At the end of this module, you will be able to understand Normal distribution, interpreting z-scores and calculating percentiles, Binomial Distribution, Mean and Standard deviation. You will also understand the Milgram Experiment.

Topics

- ✓ Probability Distributions & Regression Modeling
- ✓ Normal Distribution
- ✓ Binomial Distribution
- ✓ Linear Regression Model and Analysis

Big Data Hadoop Certification Training

Course Curriculum

About The Course

Become a Hadoop Expert by mastering MapReduce, Yarn, Pig, Hive, HBase, Oozie, Flume and Sqoop while working on industry based Use-cases and Projects. Also get an overview of Apache Spark for distributed data processing.

Module 1 : Understanding Big Data and Hadoop

Learning Objectives

In this module, you will understand Big Data, the limitations of the existing solutions for Big Data problem, how Hadoop solves the Big Data problem, the common Hadoop ecosystem components, Hadoop Architecture, HDFS, Anatomy of File Write and Read, how MapReduce Framework works

Topics

- ✓ Big Data
- ✓ Limitations and Solutions of existing Data Analytics Architecture
- ✓ Hadoop
- ✓ Hadoop Features
- ✓ Hadoop Ecosystem
- ✓ Hadoop Storage: HDFS, Hadoop Processing: MapReduce Framework
- ✓ Hadoop 2.x core components
- ✓ Hadoop Different Distributions

Module 2 : Hadoop Architecture and HDFS

Learning Objectives

In this module, you will learn the Hadoop Cluster Architecture, Important Configuration files in a Hadoop Cluster, Data Loading Techniques, how to setup single node and multi node Hadoop cluster

Topics

- ✓ Hadoop 2.x Cluster Architecture
 - Federation and High Availability
- ✓ A Typical Production Hadoop Cluster
- ✓ Multi node cluster set up Hadoop Administration
- ✓ Hadoop Cluster Modes
- ✓ Common Hadoop Shell Commands
- ✓ Hadoop 2.x Configuration Files
- ✓ Single node cluster

Module 3 : Hadoop MapReduce Framework

Learning Objectives

In this module, you will understand Hadoop MapReduce framework and the working of MapReduce on data stored in HDFS. You will understand concepts like Input Splits in MapReduce, Combiner & Partitioner and Demos on MapReduce using different data sets.

Topics

- ✓ MapReduce Use Cases
- ✓ Traditional way Vs MapReduce way
- ✓ Why MapReduce
- ✓ Hadoop 2.x MapReduce Architecture
- ✓ Hadoop 2.x MapReduce Components
- ✓ YARN MR Application Execution Flow
- ✓ YARN Workflow, Anatomy of MapReduce Program

- ✓ Hadoop 2.x MapReduce Architecture
- ✓ MapReduce: Combiner & Partitioner
- ✓ Demo on de-identifying Health Care Data set
- ✓ Demo on MapReduce. Input Splits, Relation between Input Splits and HDFS Blocks
- ✓ Demo on Weather Data set.

Module 4 : Advanced MapReduce

Learning Objectives

In this module, you will learn Advanced MapReduce concepts such as Counters, Distributed Cache, MRunit, Reduce Join, Custom Input Format, Sequence Input Format and XML parsing.

Topics

- ✓ Counters
- ✓ Distributed Cache
- ✓ MRunit
- ✓ Sequence Input Format
- ✓ Reduce Join
- ✓ Custom Input Format
- ✓ Xml file Parsing using MapReduce

Module 5 : Pig

Learning Objectives

In this module, you will learn Pig, types of use case we can use Pig, tight coupling between Pig and MapReduce, and Pig Latin scripting, PIG running modes, PIG UDF, Pig Streaming, Testing PIG Scripts. Demo on healthcare dataset.

Topics

- ✓ About Pig
- ✓ MapReduce Vs Pig
- ✓ Pig Use Cases
- ✓ Programming Structure in Pig
- ✓ Data Models in Pig
- ✓ Shell and Utility Commands
- ✓ Built In Functions (Eval Function, Load and Store Functions, Math function, String Function, Date Function, Pig UDF, Piggybank
- ✓ Parameter Substitution (PIG macros and Pig Parameter substitution)
- ✓ Testing Pig scripts with Punit
- ✓ Pig Running Modes
- ✓ Pig components
- ✓ Pig Execution
- ✓ Pig Latin Program
- ✓ Pig Data Types
- ✓ Pig Latin : Relational Operators File Loaders, Group Operator, COGROUP Operator, Joins and COGROUP, Union, Diagnostic Operators, Specialized joins in Pig
- ✓ Pig Streaming
- ✓ Aviation use case in PIG, Pig Demo on Healthcare Data set.

Module 6 : Hive

Learning Objectives

This module will help you in understanding Hive concepts, Hive Data types, loading and Querying Data in Hive, running hive scripts and Hive UDF.

Topics

- ✓ Hive Background
- ✓ Hive Use Case
- ✓ About Hive
- ✓ Hive Vs Pig

- ✓ Hive Architecture and Components
- ✓ Limitations of Hive
- ✓ Hive Data Types and Data Models
- ✓ Programming Structure in Pig
- ✓ Hive Tables(Managed Tables and External Tables)
- ✓ Querying Data
- ✓ Hive Script
- ✓ Retail use case in Hive
- ✓ Metastore in Hive
- ✓ Comparison with Traditional Database
- ✓ Partitions and Buckets
- ✓ Pig Latin Program
- ✓ Importing Data
- ✓ Managing Outputs
- ✓ Importing Data
- ✓ Hive UDF
- ✓ Hive Demo on Healthcare Data set.

Module 7 : Advanced Hive and HBase

Learning Objectives

In this module, you will understand Advanced Hive concepts such as UDF, Dynamic Partitioning, Hive indexes and views, optimizations in hive. You will also acquire in-depth knowledge of HBase, HBase Architecture, running modes and its components.

Topics

- ✓ Hive QL: Joining Tables, Dynamic Partitioning, Custom Map/Reduce Scripts
- ✓ Hive : Thrift Server, User Defined Function
- ✓ Hive Indexes and views
- ✓ Hive query optimizers
- ✓ HBase: Introduction to NoSQL Databases and HBase, HBase v/s RDBMS, HBase Components, HBase Architecture, HBase Cluster Deployment

Module 8 : Advance HBase

Learning Objectives

This module will cover Advanced HBase concepts. We will see demos on Bulk Loading, Filters. You will also learn what Zookeeper is all about, how it helps in monitoring a cluster, why HBase uses Zookeeper.

Topics

- ✓ HBase Data Model
- ✓ HBase Client API
- ✓ ZooKeeper Data Model
- ✓ Zookeeper
- ✓ Getting and Inserting Data
- ✓ HBase Shell
- ✓ Data Loading Techniques
- ✓ Zookeeper Service
- ✓ Demos on Bulk Loading
- ✓ Filters in HBase

Module 9 : Processing Distributed Data with Apache Spark

Learning Objectives

In this module you will learn Spark ecosystem and its components, how Scala is used in Spark, SparkContext. You will learn how to work in RDD in Spark. Demo will be there on running application on Spark Cluster, Comparing performance of MapReduce and Spark.

Topics

- ✓ What is Apache Spark
- ✓ Spark Components
- ✓ Spark a Polyglot
- ✓ Why Scala?
- ✓ Spark Ecosystem
- ✓ History of Spark and Spark Versions/ Releases
- ✓ What is Scala?

✓ SparkContext

✓ RDD

Module 10 : Oozie and Hadoop Project

Learning Objectives

In this module, you will understand working of multiple Hadoop ecosystem components together in a Hadoop implementation to solve Big Data problems. We will discuss multiple data sets and specifications of the project. This module will also cover Flume & Sqoop demo, Apache Oozie Workflow Scheduler for Hadoop Jobs, and Hadoop Talend integration.

Topics

✓ Scheduling with Oozie

✓ Oozie Co-ordinator

✓ Oozie Web Console

✓ PIG, Hive, and Sqoop,

✓ Hadoop Project Demo

✓ Demo on Oozie Workflow

✓ Oozie Commands

✓ Oozie for MapReduce

✓ Combine flow of MR, PIG, Hive in Oozie

✓ Hadoop Integration with Talend

Project Work

Towards the end of the course, you will be working on a live project where you will be using PIG, HIVE, HBase and MapReduce to perform Big Data analytics.

Here are the few Industry-wise Big Data case studies e.g. Finance, Retail, Media, and Aviation etc. which you can take up as your project work:

Project 1 : Analyze social bookmarking sites to find insights

Industry: Social Media

Data: It comprises of the information gathered from sites like reddit.com, stumbleupon.com etc. which are bookmarking sites and allow you to bookmark, review, rate, search various links on any topic. reddit.com, stumbleupon.com, etc. A bookmarking site allows you to bookmark, review, and rate, search various links on any topic. The data is in XML format and contains various links/posts URL, categories defining it and the ratings linked with it.

Problem Statement: Analyze the data in Hadoop Eco-system to:

1. Fetch the data into Hadoop Distributed File System and analyze it with the help of MapReduce, Pig and Hive to find the top rated links based on the user comments, likes etc.
2. Using MapReduce convert the semi-structured format (XML data) into structured format and categorize the user rating as positive and negative for each of the thousand links.
3. Push the output HDFS and then feed it into PIG, which splits the data into two parts: Category data and Ratings data.
4. Write a fancy Hive Query to analyze the data further and push the output is into relational database (RDBMS) using Sqoop.
5. Use a web server running on rails/java/ruby/python that renders the result in real time processing on a website.

Project 2 : Customer Complaints Analysis

Industry: Retail

Data: Publicly available dataset, containing a few lakh observations with attributes like: Customer ID, Payment Mode, Product Details, Complaint, Location, Status of the complaint, etc.

Problem Statement: Analyze the data in Hadoop Eco-system to:

1. Get the number of complaints filed under each products
2. Get the total number of complaints filed from a particular location
3. Get the list of complaints grouped by location which has no timely response

Project 3 : Tourism Data Analysis

Industry: Tourism

Data: The dataset comprises attributes like: City pair (Combination of from and to), Adults traveling, Seniors traveling, Children traveling, Air booking price, Car booking price, etc.

Problem Statement: Find the following insights from the data:

1. Top 20 destinations people travel most: Based on given data we can find the most popular destinations where people travel frequently, based on the specific initial number of trips booked for a particular destination
2. Top 20 locations from where most of the trips start based on booked trip count
3. Top 20 high air-revenue destinations i.e. which 20 cities generates high airline revenues for travel, so that the discount offers can be given to attract more bookings for these destinations

Project 4 : Airline Data Analysis

Industry: Aviation

Data: Publicly available dataset which contains the flight details of various airlines like: Airport id, Name of the airport, Main city served by airport, Country or territory where airport is located, Code of Airport, Decimal degrees, Hours offset from UTC, Time zone, etc.

Problem Statement: Analyze the airlines data to:

1. Find list of Airports operating in the Country
2. Find the list of Airlines having zero stops
3. List of Airlines operating with code share
4. Which country (or) territory has the highest number of Airports
5. Find the list of Active Airlines in the United States

Project 5 : Analyze Loan Dataset

Industry: Banking and Finance

Data: Publicly available dataset which contains complete details of all the loans issued, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information.

Problem Statement: Find the number of cases per location and categorize the count with respect to reason for taking loan and display the average risk score

Project 6 : Analyze Movie Ratings

Industry: Media

Data: Publicly available data from sites like rotten tomatoes, imdb, etc.

Problem Statement: Find the number of cases per location and categorize the count with respect to reason for taking loan and display the average risk score

Project 7 : Analyze YouTube data

Industry: Social Media

Data: It is about the YouTube videos and contains attributes like: Video ID, Uploader, Age, Category, Length, views, ratings, comments, etc.

Problem Statement: Identify out the top 5 categories in which the most number of videos are uploaded, the top 10 rated videos, the top 10 most viewed videos

Apart from these there are some twenty more use-cases to choose from:

Market data Analysis

Twitter Data Analysis

Apache Spark and Scala

Course Curriculum

About The Course

The Edureka Apache Spark & Scala course will enable learners to understand how Spark enables in-memory data processing and runs much faster than Hadoop MapReduce. Learners learn about RDDs, different APIs which Spark offers such as Spark Streaming, MLlib, SparkSQL, GraphX. This Edureka course is an integral part of a developer's learning path.

Module 1 : Introduction to Scala for Apache Spark

Learning Objectives

In this module, you will understand the basics of Scala that are required for programming Spark applications. You can learn about the basic constructs of Scala such as variable types, control structures, collections, and 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, Procedures, Collections in Scala-Array
- ✓ ArrayBuffer
- ✓ Map, Tuples, Lists, and more.

Module 2 : OOPS and Functional Programming in Scala

Learning Objectives

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

Topics

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

Module 3 : Introduction to Big Data and Apache Spark

Learning Objectives

In this module, you will understand what is big data, challenges associated with it and the different frameworks available. The module also includes a first-hand introduction to Spark.

Topics

- ✓ Introduction to big data
- ✓ Challenges with big data
- ✓ Batch Vs. Real Time big data analytics
- ✓ Batch Analytics - Hadoop Ecosystem Overview
- ✓ Real-time Analytics Options
- ✓ Streaming Data- Spark
- ✓ In-memory data- Spark
- ✓ What is Spark?
- ✓ Spark Ecosystem,
- ✓ Modes of Spark
- ✓ Spark installation demo
- ✓ Overview of Spark on a cluster
- ✓ Spark Standalone cluster
- ✓ Spark Web UI.

Module 4 : Spark Common Operations

Learning Objectives

In this module, you will learn how to invoke Spark Shell and use it for various common operations

Topics

- ✓ Invoking Spark Shell
- ✓ Creating the Spark Context
- ✓ Loading a file in Shell
- ✓ Performing basic Operations on files in Spark Shell
- ✓ Overview of SBT
- ✓ Building a Spark project with SBT
- ✓ Running Spark project with SBT
- ✓ Local mode
- ✓ Spark mode
- ✓ Caching overview
- ✓ Distributed Persistence

Module 5 : Playing with RDDs

Learning Objectives

In this module, you will learn one of the fundamental building blocks of Spark - RDDs and related manipulations for implementing business logics.

Topics

- ✓ RDDs
- ✓ Transformations in RDD
- ✓ Actions in RDD
- ✓ Loading data in RDD
- ✓ Saving data through RDD
- ✓ Key-Value Pair RDD
- ✓ MapReduce and Pair RDD Operations
- ✓ Spark and Hadoop Integration-HDFS
- ✓ Spark and Hadoop Integration-Yarn
- ✓ Handling Sequence Files and Partitioner

Module 6 : Spark Streaming and MLlib

Learning Objectives

In this module, you will learn about the major APIs that Spark offers. You will get an opportunity to work on Spark streaming which makes it easy to build scalable fault-tolerant streaming applications, MLlib which is Spark's machine learning library.

Topics

- ✓ Spark Streaming Architecture
- ✓ First Spark Streaming Program
- ✓ Transformations in Spark Streaming
- ✓ Fault tolerance in Spark Streaming
- ✓ Checkpointing
- ✓ Parallelism level
- ✓ Machine learning with Spark
- ✓ Data types

- ✓ Algorithms– statistics
- ✓ Classification and regression
- ✓ Clustering
- ✓ Collaborative filtering.

Module 7 : GraphX, SparkSQL and Performance Tuning in Spark

Learning Objectives

In this module, you will learn about Spark SQL that is used to process structured data with SQL queries, graph analysis with Spark, GraphX for graphs and graph-parallel computation. You will also get a chance to learn the various ways to optimize performance in Spark.

Topics

- ✓ Analyze Hive and Spark SQL architecture
- ✓ SQLContext in Spark SQL
- ✓ Working with DataFrames
- ✓ Implementing an example for Spark SQL
- ✓ Integrating hive and Spark SQL
- ✓ Support for JSON and Parquet File Formats
- ✓ Implement data visualization in Spark
- ✓ Loading of data
- ✓ Hive queries through Spark
- ✓ Testing tips in Scala
- ✓ Performance tuning tips in Spark
- ✓ Shared variables: Broadcast Variables
- ✓ Shared Variables: Accumulators.

Module 8 : A complete project on Apache Spark

Learning Objectives

In this module, you will get an opportunity to work on a live Spark project where you can implement the learnings from previous modules hands-on, and solve a real-time use case.

Problem Statement

Design a system to replay the real time replay of transactions in HDFS using Spark.

Technologies used

- ✓ Spark Streaming
- ✓ HDFS (for storage)
- ✓ Kafka (for messaging)
- ✓ Core Spark API (for aggregation)

Project Work

At the end of the course, you will be working on a live project:

Project 1 : Analyzing the telecom data to minimize churn out

Industry: Telecom

Data: It is in the form of a CDR (Call Details Record) file.

Problem Statement: Find out top 10 customers facing frequent call drops in Roaming.

This is a significant report used frequently by telecom companies to prevent customer churn out. This is done by retargeting the customers through different follow-ups and at the same time contacting their roaming partners to improve the connectivity issues in specific areas.

Deep Learning using TensorFlow

Course Curriculum

About The Course

In this Training, you will be able to learn the basic concepts of TensorFlow, the main functions, operations and the execution pipeline. Starting with a simple "Hello Word" example, throughout the course you will be able to see how TensorFlow can be used in curve fitting, regression, classification and minimization of error functions. This concept is then explored in the Deep Learning world. You will evaluate the common, and not so common, deep neural networks and see how these can be exploited in the real world with complex raw data using TensorFlow. In addition, you will learn how to apply TensorFlow for backpropagation to tune the weights and biases while the Neural Networks are being trained. Finally, the course covers different types of Deep Architectures, such as Convolutional Networks, Recurrent Networks and Autoencoders. Delve into neural networks, implement Deep Learning algorithms, and explore layers of data abstraction with the help of this Deep Learning using TensorFlow Certification Training.

Module 1 : Introduction to Deep Learning

Learning Objectives

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

1. Discuss the revolution of Artificial Intelligence

2. Discuss the limitations of Machine Learning
3. List the advantages of Deep Learning over Machine Learning
4. Discuss Real-life use cases of Deep Learning
5. Understand the Scenarios where Deep Learning is applicable
6. Discuss relevant topics of Linear Algebra and Statistics
7. Discuss Machine learning algorithms
8. Define Reinforcement Learning
9. Discuss model parameters and optimization techniques

Topics

- ✓ Deep Learning: A revolution in Artificial Intelligence
- ✓ Advantage of Deep Learning over Machine learning
- ✓ Real-Life use cases of Deep Learning
- ✓ The Math behind Machine Learning: Linear Algebra
 - Scalars
 - Vectors
 - Matrices
 - Tensors
 - Hyperplanes
- ✓ Review of Machine Learning Algorithms
 - Regression
 - Classification
 - Clustering
- ✓ Limitations of Machine Learning
- ✓ Discuss the idea behind Deep Learning
- ✓ 3 Reasons to go Deep
- ✓ Scenarios where Deep Learning is applicable
- ✓ The Math Behind Machine Learning: Statistics
 - Probability
 - Conditional Probabilities
 - Posterior Probability
 - Distributions
 - Samples vs Population
 - Resampling Methods
 - Selection Bias
 - Likelihood

- ✓ Reinforcement Learning
- ✓ Optimization
- ✓ Underfitting and Overfitting
- ✓ Convex Optimization

Module 2 : Fundamentals of Neural Networks

Learning Objectives

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

1. Define Neural Networks
2. Discuss the Training Techniques of Neural Networks
3. List Different Activation and Loss Functions
4. Discuss the Different parameters of Neural Networks

Topics

- ✓ Defining Neural Networks
- ✓ The Perceptron
- ✓ Training Neural Networks
- ✓ Gradient Descent
- ✓ Quasi-Newton Optimization Methods
- ✓ Activation Functions
 - Linear
 - Sigmoid
 - Tanh
 - Hard Tanh
 - Softmax
 - Rectified Linear
- ✓ The Biological Neuron
- ✓ Multi-Layer Feed-Forward Networks
- ✓ Backpropagation Learning
- ✓ Stochastic Gradient Descent
- ✓ Generative vs Discriminative Models
 - Loss Functions
 - Loss Function Notation
 - Loss Functions for Regression
 - Loss Functions for Classification
 - Loss Functions for Reconstruction
 - Hyperparameters

- Learning Rate
- Momentum
- Regularization
- Sparsity

Module 3 : Fundamentals of Deep Networks

Learning Objectives

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

1. Define Deep Learning
2. Discuss the Architectural Principals of Deep Networks
3. List Different parameters of Deep Networks
4. Discuss the Building Blocks of Deep Networks
5. Discuss how reinforcement learning is used in Deep Networks

Topics

- | | |
|-------------------------------------------------------|----------------------------------------------|
| ✓ Defining Deep Learning | ✓ Layers |
| ✓ Defining Deep Networks | ✓ Activation Functions - Sigmoid, Tanh, ReLU |
| ✓ Common Architectural Principals of Deep Networks | ✓ Loss Functions |
| ✓ Reinforcement Learning application in Deep Networks | ✓ Optimization Algorithms |
| ✓ Parameters | ✓ Hyperparameters |
| | ✓ Summary |

Module 4 : Introduction to TensorFlow

Learning Objectives

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

1. Define TensorFlow
2. Illustrate how TensorFlow works
3. Discuss the Functionalities of TensorFlow
4. Illustrate different ways to install TensorFlow
5. Write and Run programs on TensorFlow

Topics

- | | |
|--------------------------------------|-------------------------------------------------------|
| ✓ What is TensorFlow? | ✓ HelloWorld with TensorFlow |
| ✓ Use of TensorFlow in Deep Learning | ✓ Running a Machine learning algorithms on TensorFlow |
| ✓ Working of TensorFlow | |
| ✓ How to install Tensorflow | |

Module 5 : Convolutional Neural Networks (CNN)

Learning Objectives

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

1. Define CNNs
2. Discuss the Applications of CNN
3. Explain the Architecture of a CNN
4. List Convolution and Pooling Layers in CNN
5. Illustrate CNN
6. Discuss Fine-tuning and Transfer Learning of CNNs

Topics

- ✓ Introduction to CNNs
- ✓ CNNs Application
- ✓ Architecture of a CNN
- ✓ Understanding and Visualizing a CNN
- ✓ Convolution and Pooling layers in a CNN
- ✓ Transfer Learning and Fine-tuning Convolutional Neural Networks

Module 6 : Recurrent Neural Networks (RNN)

Learning Objectives

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

1. Define RNN
2. Discuss the Applications of RNN
3. Illustrate how RNN is trained
4. Discuss Long Short-Term memory(LSTM)
5. Explain Recursive Neural Tensor Network Theory
6. Illustrate the working of Neural Network Model

Topics

- ✓ Intro 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

Module 7 : Restricted Boltzmann Machine(RBM) and Autoencoders

Learning Objectives

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

1. Define RBM
2. Discuss the Applications of RBM
3. Illustrate Collaborative Filtering using RBM
4. Define Autoencoders
5. Explain Deep Belief Networks

Topics

- | | |
|------------------------------------|------------------------------|
| ✓ Restricted Boltzmann Machine | ✓ Autoencoders applications |
| ✓ Applications of RBM | ✓ Understanding Autoencoders |
| ✓ Collaborative Filtering with RBM | ✓ Variational Autoencoders |
| ✓ Introduction to Autoencoders | ✓ Deep Belief Network |