Description

The course content is designed in a way that is simple to follow and understand; the course is expressive, exhaustive, practical with *live* coding, replete with quizzes, and rich with state-of-the-art and up-to-date knowledge of this field.

What You will Learn

- Introduction and importance of this course in this day and age
- Approach all essential concepts from the beginning
- Clear unfolding of concepts with examples in Python, Scrapy, Scala, PySpark, and MongoDB
- All theoretical explanations followed by practical implementations
- Data scraping and data mining for beginners to pro with Python
- Master Big Data with Scala and Spark
- Master Big Data with PySpark and AWS
- Mastering MongoDB for beginners
- Build your own AI applications

I. Scala

It's true that Scala is not among the most-loved coding languages but don't let this minor discomfort bother you. Scala is doubtless one of the most in-demand skills for data scientists and data engineers. And the reason for this is not far to seek: The supply of professionals with Scala skills is a long way from catching up with the demand.

The well-thought-out quizzes and mini-projects in this course will cover all the important aspects and it will make your Scala learning journey that much easier. This course includes an overview of Hadoop and Spark with a hands-on project with Scala Spark. Right through the course, every theoretical explanation is followed by practical implementation.

This course is designed to reflect the most in-demand Scala skills that you will start using right away at the workplace. The six mini-projects and one Scala Spark project included are a vital component of this course. These projects present you with a hands-on opportunity to experiment for yourself with trial and error. You get a chance to learn from the mistakes you commit.

Importantly, it is easy to understand the potential gaps that might exist between theory and practice.

Scala, a power-packed language, has the capability to leverage most of the functions in Python, such as designing machine learning models. You can use this high-level language for an assortment of applications, from web apps to machine learning.

II. PySpark and AWS

The hottest buzzwords in the Big Data analytics industry are Python and Apache Spark. PySpark supports the collaboration of Python and Apache Spark. In this course, you will start right from the basics and proceed to the advanced levels of data analysis. From cleaning data to building features and implementing machine learning (ML) models, you will learn how to execute end-to-end workflows using PySpark.

Right through the course, you will be using PySpark for performing data analysis. You will explore Spark RDDs, Dataframes, and a bit of Spark SQL queries. You will also explore the transformations and actions that can be performed on the data using Spark RDDs and dataframes. You will explore the ecosystem of Spark and Hadoop and their underlying architecture. You will use the Databricks environment for running the Spark scripts and explore it as well.

Finally, you will have a taste of Spark with AWS cloud. You will see how we can leverage AWS storages, databases, computations, and how Spark can communicate with different AWS services and get its required data.

As this course is a detailed compilation of all the basics, it will motivate you to make quick progress and experience much more than what you have learned. At the end of each concept, you will be assigned homework/tasks/activities/quizzes along with solutions. This is to evaluate and promote your learning based on the previous concepts and methods you have learned. Most of these activities will be coding-based, as the aim is to get you up and running with implementations.

III. Data Scraping and Data Mining from Beginner to Professional

Data scraping is the technique of extracting data from the Internet. Data scraping is used to get the data available on different websites and APIs. This also involves automating the web flows to extract data from different webpages.

This course is designed for beginners. We will spend sufficient time in laying a solid groundwork for newbies. Then, we will go deep gradually with a lot of practical implementations where every step will be explained in detail.

As this course is essentially a compilation of all the basics, you will move ahead at a steady rate. You will experience more than what you have learned. At the end of every concept, we will be assigning you homework/assignments/activities/quizzes along with solutions. They will assess (further build) your learning based on the previous data scraping and data mining concepts and methods. Most of these activities are designed to get you up and running with implementations.

The four hands-on projects included in this course are the most important part of this course. These projects allow you to experiment for yourself with trial and error. You will learn from your mistakes. Importantly, you will understand the potential gaps that may exist between theory and practice.

Data scraping is undoubtedly a rewarding career that allows you to solve some of the most interesting real-world problems. You will be rewarded with a fabulous salary package too. With a core understanding of data scraping, you can fine-tune your workplace skills and ensure emerging career growth.

IV. MongoDB

In this course, we will go through the basics of MongoDB. We will be using MongoDB to develop the understanding of NoSQL databases. We will explore the basic Create, Read, Update, and Delete operations in MongoDB. We will explore in detail the MongoDB query operators and project operators. Following that, you will learn about MongoDB update operators. In the end, we will move to explore MongoDB with Node and Python. We will wind up this course with two projects, consisting of MongoDB with Django, in which we will develop a CRUD-based application using Django and MongoDB and then we will implement an ETL pipeline using PySpark to dump the data in MongoDB.

This course is designed for beginners. We will spend enough time to make a solid ground for newbies and we will go deep gradually with a lot of practical implementations where every step will be explained in detail.

As this course is a compilation of all the basics, it will encourage you to move ahead and experience more than what you have learned. By the end of every concept, we will be assigning you homework/tasks/activities/quizzes along with solutions that will evaluate (further build)

your learning based on the previous concepts and methods. Several of these activities will be coding-based to get you up and running with implementations.

With the increase of data, there is a need to manage that; and not only manage it but also get the useful data and insights out of it for business analytics and correct decision making; for this, companies are actively looking for big data engineers. The major issue with big data is that it is so humongous that using regular data analysis techniques to analyze it is not possible. Also, due to continuously increasing data sources such as IoT, SQL databases, NoSQL databases, social media platforms, point of sales, and streaming data, it is hard to even manage all this data through conventional methods; performing analytics on it is, as I just mentioned, way beyond these methods. So, we need new techniques and platforms for not only managing this data but also performing analysis on it; MongoDB supports all of this. We will understand and learn using MongoDB, which, in a nutshell, is a NoSQL database. All these skills are highly in demand.

So, without any further delay, let us get started with the course and embrace yourself with the knowledge that waits for you.

1. **Scope of Scala**:

    - Understanding the variables in data types in Scala
    - Understanding the flow controls in Scala and different ways for controlling the flow
    - Understanding the functions and their usage in Scala
    - Understanding the classes and their usage in Scala
    - Understanding the data structures, namely: lists, lists buffer, maps, sets, and stack
    - Understanding Hadoop
    - Understanding the working of Spark
    - Understanding the difference between Spark RDDs and Spark DFs
    - Understanding Map Reduce
    - Understanding the ETL pipeline from AWS S3 to AWS RDS using Spark

1. **Scope of PySpark**:
    - Spark/Hadoop applications, ecosystem, and architecture
    - PySpark RDDs
    - PySpark RDD transformations
    - PySpark RDD actions
    - PySpark DataFrames

Commented [TF1]: Understanding ETL...?

Commented [KS2R1]: yes, please update it

- PySpark DataFrames transformations
- PySpark DataFrames actions
- Collaborative filtering in PySpark
- Spark Streaming
- ETL Pipeline
- CDC and replication on Going

2. **Scope of Data Scraping, Data Mining**:
    - Internet browser execution and communication with the server
    - Request/response to and from the server
    - Synchronous and asynchronous
    - Parsing data in response from the server
    - Difference between synchronous and asynchronous requests
    - Introductions to tools for data scraping: Requests, BS4, Scrapy and Selenium
    - Explanation of different concepts such as Python Requests Module, BS4 parsers functions, Scrapy for writing the spiders for crawling websites and extracting data, Selenium for understanding the automation and control of web flows, and more
3. **Scope of MongoDB**:
    - Understanding MongoDB CRUD, query operators, projection operators, update operators
    - Creating MongoDB cluster on Atlas
    - Understanding MongoDB with Node
    - Performing CRUD operation with Node in MongoDB Atlas
    - Understanding MongoDB with Python
    - Performing CRUD operation with Python in MongoDB Atlas
    - Understanding MongoDB with Django
    - Performing CRUD operation with Django in MongoDB Atlas
    - Building APIs for CRUD operations in MongoDB through Django
    - Understanding MongoDB with PySpark

After completing this information-packed course successfully, you will be able to perform the following:

● Implement any project from scratch that requires data scraping, data mining, Scala, PySpark, AWS, and MongoDB knowledge

● Relate the concepts and practical aspects of learned technologies with real-world problems

Commented [TF3]: Can this be rephrased as a single sentence?

Commented [KS4R3]: made it 2 diff pointers

Commented [TF5]: Is this okay?

Commented [KS6R5]: yes

● Gather data from websites in the smartest way

Who this course is for:

● People who are absolute beginners

● People who want to make smart solutions

● People who want to learn with real data

● People who love to learn theory and then implement it practically

● Data scientists, machine learning experts, and drop shippers

Requirements

- Basic understanding of HTML tags, Python, SQL, and Node JS.
- No prior knowledge of data scraping and Scala is needed. You start right from the basics and then gradually build your knowledge of the subject.
- Basic understanding of programming.
- A willingness to learn and practice.
- We teach by practical implementations, so practice is a must thing to do.