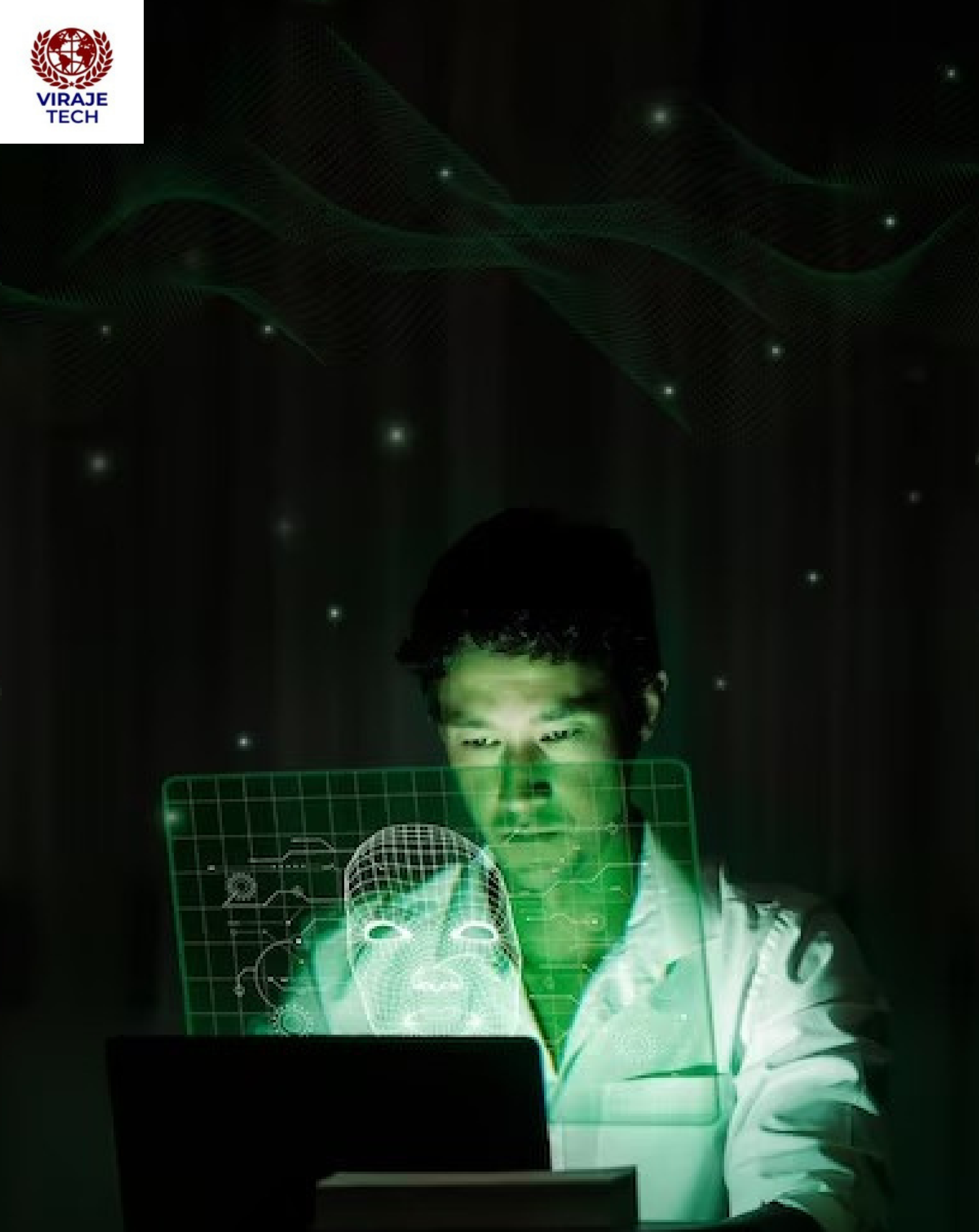
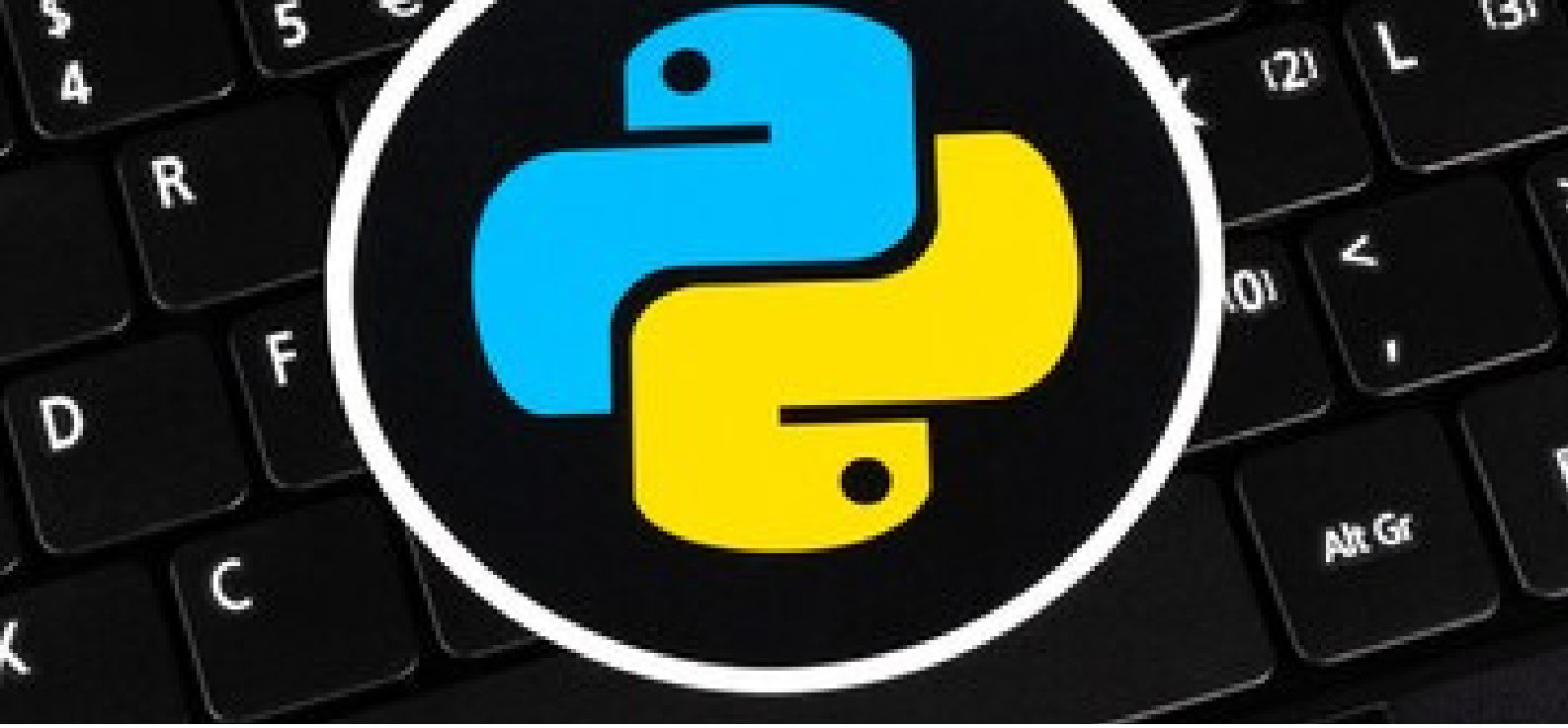
VIRAJE TECH



# BIG DATA HADOOP


- Access to Interview Opportunities with Top Companies
- Industry-Relevant Curriculum Designed and Taught by Industry Experts
- Hands on Project and Industry Specific Tools
- Dedicated CareerSupport and Interview Preparation
- Post Graduate Certificatefrom Great Lakes Executive Learning



This Apache Spark course is essential for anyone seeking proficiency in big data processing and analytics. As an open-source distributed computing system, Spark is renowned for its efficiency and scalability. The course offers a hands-on approach to mastering core concepts like RDDs and DataFrames, enabling participants to harness Spark's in-memory processing capabilities. By providing practical experience in building fault-tolerant data processing applications, using Spark's APIs in languages such as Scala or Python, the course empowers learners to efficiently analyze vast datasets. Additionally, participants will delve into Spark's ecosystem, exploring modules like Spark SQL, Spark Streaming, and MLlib, fostering a comprehensive skill set for diverse data processing needs.

The Program helps you do grow and bloom in Industry and developed by best-in-class industry experts.It offers a blend of online learning with live and recorded lectures along with access to dedicated career support and rewarding job opportunities.

## LEARN ONLINE ANYTIME, ANYWHERE

Learn from live masterclasses by top industryleaders and online lab sessions every week, along with 100+ hours of learning content.

## WEEKLY ONLINE MENTORSHIPFROM EXPERTS

Get assistance on projects and reinforce the concepts you learn through weekly mentorship sessions.

## NETWORK WITH LIKE-MINDED PEERS

Interact with peers from diverse backgroundsand grow your professional network.

## DEDICATED PROGRAM SUPPORT

Access dedicated support on your learning journey and resolve for all your queries withhelp from a dedicated Program Manager.

A fresh graduate or a working professional looking to up-skill andbuild a career.

# LEARNING PLAN

## APACHE SPARK

## Module1:  Introduction to Big Data and Hadoop

1.1.What is Big Data?

- 1.1.1. Definition and Characteristics

- 1.1.2. Sources of Big Data

1.2.Challenges of Big Data

- 1.2.1. Volume, Velocity, Variety, and Veracity

- 1.2.2. Privacy and Security Concerns

1.3.Introduction to Hadoop

- 1.3.1. Origins of Hadoop

- 1.3.2. Hadoop's Role in Big Data

## Module 2:Hadoop Fundamentals

2.1.Hadoop Distributed File System (HDFS)

- 2.1.1. Architecture and Concepts

- 2.1.2. HDFS Commands and Operations

2.2.MapReduce

- 2.2.1. MapReduce Basics

- 2.2.2. MapReduce Workflow

2.3.YARN (Yet Another Resource Negotiator)

- 2.3.1. Introduction and Role in Hadoop

# Module 3:Hadoop Ecosystem

3.1.Hive

- 3.1.1. Introduction to Hive

- 3.1.2. HiveQL (Hive Query Language)

3.2.Pig

- 3.2.1. Introduction to Pig

- 3.2.2. Writing Pig Latin Scripts

3.3.HBase

- 3.3.1. Introduction to HBase

- 3.3.2. HBase Data Model

# Module 4:Hadoop Installation and Configuration

4.1.Hadoop Installation

- 4.1.1. Single-Node Cluster Setup

- 4.1.2. Multi-Node Cluster Setup

4.2.Configuration Files

- 4.2.1. Core-site.xml and HDFS-site.xml

- 4.2.2. MapReduce Configuration

4.3.Cluster Management

- 4.3.1. Namenode and Datanode Management

- 4.3.2. ResourceManager and NodeManager

# Module 5: Data Ingestion and Processing with Hadoop

5.1.Importing Data into HDFS

- 5.1.1. Hadoop Copy Commands

- 5.1.2. Data Loading Techniques

5.2.Data Processing with MapReduce

- 5.2.1. Writing MapReduce Programs

- 5.2.2. Running MapReduce Jobs

## Module 6: Hadoop Data Storage and Management

6.1.Hive for Data Warehousing

- 6.1.1. Creating Hive Tables

- 6.1.2. Running Hive Queries

6.2.Pig for Data Transformation

- 6.2.1. Data Transformation with Pig

- 6.2.2. Complex Data Processing

## Module 8:Hadoop Administration and Monitoring

7.1.Cluster Monitoring and Maintenance

- 7.1.1. Monitoring Cluster Health

- 7.1.2. Cluster Upgrades and Patches

7.2.Backup and Recovery

- 7.2.1. Backup Strategies

- 7.2.2. Data Recovery Procedures

-

## Module 8:Advanced Hadoop Topics

8.1.Hadoop Security

- 8.1.1. Authentication and Authorization

- 8.1.2. Kerberos Integration

8.2.Hadoop Ecosystem Extensions

- 8.2.1. Introduction to Spark

- 8.2.2. Integrating Spark with Hadoop

## Module 9:Hadoop Use Cases and Best Practices

9.1.Real-World Applications

- 9.1.1. Big Data in Industry

- 9.1.2. Case Studies

9.2.Best Practices

- 9.2.1. Performance Tuning

- 9.2.2. Scalability Strategies

## Module10: Final Project and Course Review

10.1.Project Proposal and Planning

- 10.1.1. Identifying a Real-World Problem

- 10.1.2. Designing a Hadoop Solution

10.2.Implementation and Presentation

- 10.2.1. Building and Deploying the Solution

- 10.2.2. Final Project Presentation

## READY TO ADVANCE YOUR CAREER?

**Aboutus:**https://youtu.be/TY0Bqj1F21w

**app-**https://play.google.com/store/apps/details?id=com.livecourses.virajetech
**youtube-**https://www.youtube.com/@virajetechlive1596/videos
**whatsapp group link -**
https://chat.wG2J3zSeX3eZ2Hz0nDu18UF