BIG DATA ENGINEER MASTER'S PROGRAM

In collaboration with IBM

www.simplilearn.com

Contents

About the Course

This Big Data Engineer Master's Program, in collaboration with IBM, provides training on the competitive skills required for a rewarding career in data engineering. You'll learn to master the Hadoop Big Data framework, leverage the functionality of Apache Spark with Python, simplify data lines with Apache Kafka, and use the open source database management tool MongoDB to store data in Big Data environments.

Key Features

- Industry-recognized certifications from IBM and Simplilearn
- Real-life projects providing hands-on industry training
- 30+ in-demand skills
- Lifetime access to self-paced learning and class recordings
- $1,200 worth of IBM cloud credits

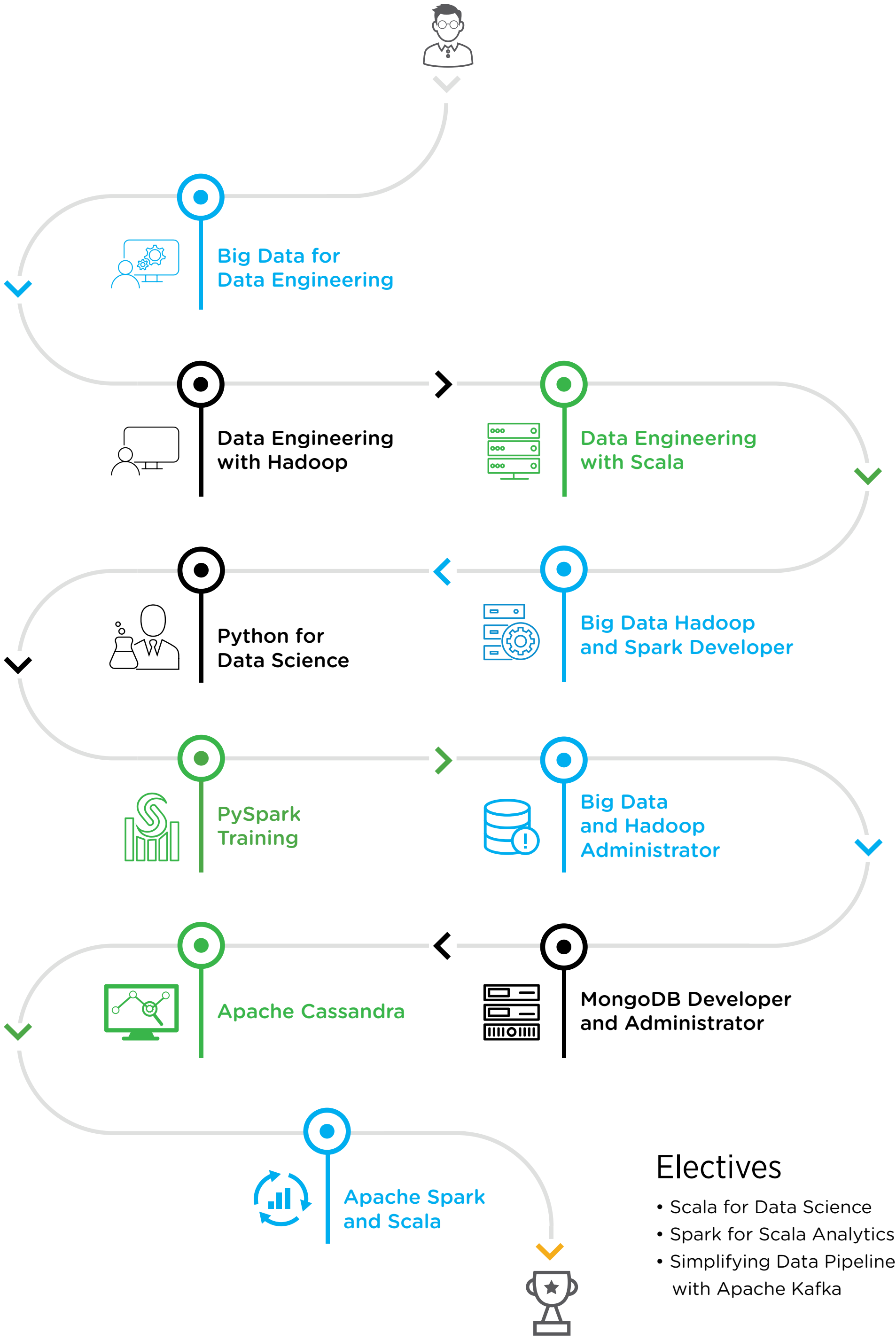
About IBM and Simplilearn collaboration

A joint partnership with Simplilearn and IBM introduces students to an integrated Blended Learning, making them an expert in Data Engineering. The program, in collaboration with IBM, will make students industry-ready for Data Engineer job roles. IBM is a leading cognitive solutions and cloud platform company, headquartered in Armonk, New York, offering a plethora of technology and consulting services. Each year, IBM invests $6 billion in research and development and has achieved five Nobel Prizes, nine US National Medals of Technology, five US National Medals of Science, six Turing Awards, and 10 Inductions in US Inventors Hall of Fame.

About Simplilearn

Simplilearn is a leader in digital skills training, focused on the emerging technologies that are transforming our world. Our blended learning approach drives learner engagement and backed by the industry's highest completion rates. Partnering with professionals and companies, we identify their unique needs and provide outcome-centric solutions to help them achieve their professional goals.

Learning Path - Data Engineer



Big Data for Data Engineering

Data Engineering with Hadoop

Data Engineering with Scala

Big Data Hadoop and Spark Developer

Python for Data Science

PySpark Training

Big Data and Hadoop Administrator

MongoDB Developer and Administrator

Apache Cassandra

Apache Spark and Scala

Electives

- Scala for Data Science
- Spark for Scala Analytics
- Simplifying Data Pipelines with Apache Kafka

Big Data Engineer Master's Program Outcomes

Gain an in-depth understanding of the flexible and versatile frameworks on the Hadoop ecosystem, such as Pig, Hive, Impala, HBase, Sqoop, Flume and Yarn.

Master tools and skills including Data Model Creation, Database Interfaces, Advanced Architecture, Spark, Scala, RDD, SparkSQL, Spark Streaming, Spark ML, GraphX, Sqoop, Flume, Pig, Hive, Impala and Kafka Architecture.

Understand how to model data, perform ingestion, replicate data, and shard data using a NoSQL database management system MongoDB.

Gain expertise in creating and maintaining analytics infrastructure and own the development, deployment, maintenance, and monitoring of architecture components.

Achieve insights on how to improve business productivity by processing Big Data on platforms that can handle its volume, velocity, variety, and veracity

Learn how Kafka is used in the real world, including its architecture and components, get hands-on experience connecting Kafka to Spark, and work with Kafka Connect

Become proficient with the fundamentals of the Scala language, its tooling, and the development process

Who Should Enroll in this Program?

A Big Data Engineer builds and maintains data structures and architectures for data ingestion, processing, and deployment for large-scale data-intensive applications. It's a promising career for both new and experienced professionals with a passion for data, including:

- ✓ IT professionals
- ✓ Banking and finance professionals
- ✓ Database administrators
- ✓ Beginners in the data engineering domain
- ✓ Students in UG/ PG programs

Big Data for Data Engineering

This introductory course from IBM will teach you the basic concepts and terminologies of Big Data, and its real-life applications across multiple industries. You will gain insights on how to improve business productivity by processing large volumes of data and extract valuable information from them.

Key Learning Objectives

- ✓ Understand what Big Data is, sources of Big Data, and real-life examples
- ✓ Learn about the key difference between Big Data and Data Science
- ✓ Master how to use Big Data for operational analysis and better customer service
- ✓ Know the Ecosystem of Big Data and Hadoop framework

Course curriculum

- ✓ Lesson 1 - What is Big Data?
- ✓ Lesson 2 - Big Data: Beyond the Hype
- ✓ Lesson 3 - Big Data and Data Science
- ✓ Lesson 4 - Use Cases
- ✓ Lesson 5 - Processing Big Data

Data Engineering with Hadoop

Apache Hadoop is one of the most in-demand technologies for analyzing Big Data. This introductory Hadoop course by IBM will give you an overview of what Hadoop is and its components, such as MapReduce and HDFS. Additionally, this course will teach you to explore with large data sets and use Hadoop's method of distributed processing.

Key Learning Objectives

- Understand Hadoop's architecture and primary components, such as MapReduce and Hadoop Distributed File System (HDFS)
- Add and remove nodes from Hadoop clusters, check the available disk space on each node, and modify configuration parameters
- Learn about Apache projects that are part of the Hadoop ecosystem, including Pig, Hive, HBase, ZooKeeper, Oozie, Sqoop, Flume, and more.

Course curriculum

- Lesson 1 - Introduction to Hadoop
- Lesson 2 -Hadoop Architecture
- Lesson 3 -Hadoop Administration
- Lesson 4 -Hadoop Components

Data Engineering with Scala

Kickstart your learning of Scala with this introductory course and familiarize yourself with Scala programming. Carefully crafted by IBM, upon completion of this course you will be able to write your Scala codes, perform Big Data analysis using Scala , and create your own Scala projects.

Key Learning Objectives

- Create your own Scala Project
- Understand basic object-oriented programming methodologies in Scala
- Work with data in Scala such as pattern matching, applying synthetic methods, handling options, failures, and futures

Course curriculum

- Lesson 1 - Introduction
- Lesson 2 - Basic Object Oriented Programming
- Lesson 3 - Case Objects and Classes
- Lesson 4 - Collections
- Lesson 5 - Idiomatic Scala

Big Data Hadoop and Spark Developer

Simplilearn's Big Data Hadoop Training Course helps you master Big Data and Hadoop Ecosystem tools, such as HDFS, YARN, MapReduce, Hive, Impala, Pig, HBase, Spark, Flume, Sqoop, Hadoop Frameworks, and more concepts of Big Data processing life cycle. Throughout this online instructor-led Hadoop Training, you will be working on real-time projects on Retail, Tourism, Finance, etc. This Big Data Course also prepares you for Cloudera's CCA175 Big Data certification.

Key Learning Objectives

- Learn how to navigate the Hadoop Ecosystem and understand how to optimize its use
- Ingest data using Sqoop, Flume, and Kafka
- Implement partitioning, bucketing, and indexing in Hive
- Work with RDD in Apache Spark
- Process real-time streaming data
- Perform DataFrame operations in Spark using SQL queries
- Implement User-Defined Functions (UDF) and User-Defined Attribute Functions (UDAF) in Spark

Course curriculum

- Lesson 1 - Introduction to Bigdata and Hadoop
- Lesson 2 - Hadoop Architecture Distributed Storage (HDFS) and YARN

- ✓ Lesson 3 - Data Ingestion into Big Data Systems and ETL
- ✓ Lesson 4 - Distributed Processing MapReduce Framework and Pig
- ✓ Lesson 5 - Apache Hive
- ✓ Lesson 6 - NoSQL Databases HBase
- ✓ Lesson 7 - Basics of Functional Programming and Scala
- ✓ Lesson 8 - Apache Spark Next-Generation Big Data Framework
- ✓ Lesson 9 - Spark Core Processing RDD
- ✓ Lesson 10 - Spark SQL Processing DataFrames
- ✓ Lesson 11 - Spark MLLib Modelling BigData with Spark
- ✓ Lesson 12 - Stream Processing Frameworks and Spark Streaming
- ✓ Lesson 13 -Spark GraphX

Python for Data Science

Kickstart your learning of Python for Data Science with this introductory course and familiarize yourself with programming. Carefully crafted by IBM, upon completion of this course you will be able to write your Python scripts, perform fundamental hands-on data analysis using the Jupyter-based lab environment, and create your own Data Science projects using IBM Watson.

Key Learning Objectives

- Write your first Python program by implementing concepts of variables, strings, functions, loops, conditions
- Understand the nuances of lists, sets, dictionaries, conditions and branching, objects and classes
- Work with data in Python such as reading and writing files, loading, working, and saving data with Pandas

Course curriculum

- Lesson 1 - Python Basics
- Lesson 2 - Python Data Structures
- Lesson 3 - Python Programming Fundamentals
- Lesson 4 - Working with Data in Python
- Lesson 5 - Working with NumPy Arrays

Pyspark Training

Pyspark Training will provide an in-depth overview of Apache Spark, the open-source query engine for processing large datasets, and how to integrate it with Python using the PySpark interface. The course will show you how to build and implement data-intensive applications as you dive into the world of high-performance machine learning leveraging Spark RDD, Spark SQL, Spark MLlib, Spark Streaming, HDFS, Sqoop, Flume, Spark GraphX, and Kafka.

Key Learning Objectives

- Understand how to leverage the functionality of Python as you deploy it in the Spark ecosystem
- Master Apache Spark architecture and how to set up a Python environment for Spark
- Learn about various techniques for collecting data, RDDs and contrast them with DataFrames, how to read data from files and HDFS, and how to work with schemas
- Obtain a comprehensive knowledge of various tools that fall under the Spark ecosystem such as Spark SQL, Spark MlLib, Sqoop, Kafka, Flume and Spark Streaming
- Create and explore various APIs to work with Spark DataFrames, and learn how to aggregate, transform, filter, and sort data with DataFrames.

Course curriculum

- ✓ Lesson 1 - A brief primer on Pyspark
- ✓ Lesson 02 - Resilient Distributed Datasets
- ✓ Lesson 03 - Resilient Distributed Datasets and actions
- ✓ Lesson 04 - DataFrames and Transformations
- ✓ Lesson 05 - Data Processing with Spark DataFrames

Big Data and Hadoop Administrator

This Big Data and Hadoop Administrator training course will furnish you with the aptitudes and methodologies necessary to excel in the Big Data Analytics industry. With this Hadoop Admin training, you'll learn to work with the adaptable, versatile frameworks based on the Apache Hadoop ecosystem, including Hadoop installation and configuration, cluster management with Sqoop, Flume, Pig, Hive, Impala, and Cloudera. You'll learn Big Data implementations that have security, speed, and scale..

Key Learning Objectives

- ✓ Understand the fundamentals and characteristics of Big Data and various scalability options available to help manage huge quantities of data
- ✓ Master the concepts of the Hadoop framework, including architecture, Hadoop distributed file system, and deployment of Hadoop clusters using core or vendor-specific distributions
- ✓ Use Cloudera manager for setup, deployment, maintenance, and monitoring of Hadoop clusters
- ✓ Work with Hadoop clients, nodes for clients and web interfaces like HUE to work with Hadoop Cluster
- ✓ Use cluster planning and tools for data ingestion into Hadoop clusters, and cluster monitoring activities
- ✓ Understand security implementation to secure data and clusters

Course curriculum

- Lesson 1 - Big Data and Hadoop Introduction
- Lesson 2 - Hadoop Distributed File System (HDFS)
- Lesson 3 - Hadoop Cluster Setup and Working
- Lesson 4 - Hadoop Configurations and Daemon Logs
- Lesson 5 - Hadoop Cluster Maintenance and Administration
- Lesson 6 - Hadoop Computational Frameworks
- Lesson 7 - Scheduling: Managing Resources
- Lesson 8 - Hadoop Cluster Planning
- Lesson 9 - Hadoop Clients and Hue Interface
- Lesson 10 - Data Ingestion in Hadoop Cluster
- Lesson 11 - Hadoop Ecosystem ComponentsServices
- Lesson 12 - Hadoop Security
- Lesson 13 - Hadoop Cluster Monitoring

STEP
1
2
3
4
5
6
7
8
9
10

MongoDB Developer and Administrator

Become an expert MongoDB developer and administrator by gaining an in-depth knowledge of NoSQL and mastering skills of data modeling, ingestion, query, sharding, and data replication. The course includes industry-based projects in e-learning and telecom domains. It is best suited for database administrators, software developers, system administrators, and analytics professionals.

Key Learning Objectives

- Develop expertise in writing Java and NodeJS applications using MongoDB
- Master the skills of Replication and Sharding of data in MongoDB to optimize read/write performance
- Perform installation, configuration, and maintenance of MongoDB environment
- Get hands-on experience in creating and managing different types of indexes in MongoDB for query execution
- Proficiently store unstructured data in MongoDB
- Develop skill sets in processing huge amounts of data using MongoDB tools
- Gain proficiency in MongoDB configuration, backup methods as well as monitoring and operational strategies
- Acquire an in-depth understanding of managing DB Notes, Replica set & Master-Slave concepts

Course curriculum

- ✓ Lesson 1 - Introduction to NoSQL databases
- ✓ Lesson 2 - MongoDB: A Database for the Modern Web
- ✓ Lesson 3 - CRUD Operations in MongoDB
- ✓ Lesson 4 - Indexing and Aggregation
- ✓ Lesson 5 - Replication and Sharding
- ✓ Lesson 6 - Developing Java and Node JS Application with MongoDB
- ✓ Lesson 7 - Administration of MongoDB Cluster Operations

Apache Cassandra

This Apache Cassandra certification training will develop your expertise in working with high-volume Cassandra database management system as part of the Big Data Hadoop framework. With this Cassandra training, you will learn Cassandra concepts, features, architecture and data model, and how to install, configure and monitor open-source databases. The Casandra course is ideal for software developers and analytics professionals who wish to further their careers in the Big Data field.

Key Learning Objectives

- Describe the need for Big Data and NoSQL
- Explain the fundamental concepts of Cassandra and its architecture
- Describe the architecture of Cassandra
- Demonstrate data model creation in Cassandra
- Use Cassandra database interfaces
- Demonstrate Cassandra database configuration

Course curriculum

- Lesson 1 - Introduction to Big Data and NoSQL Databases
- Lesson 2 - Introduction to Cassandra
- Lesson 3 - Architecture of Cassandra
- Lesson 4 - Installation and Configuration of Cassandra
- Lesson 5 - Cassandra Data Model
- Lesson 6 - Cassandra Interfaces
- Lesson 7 - Advanced Architecture and Cluster Management
- Lesson 8 - Hadoop Ecosystem around Cassandra

Apache Spark and Scala

This Apache Spark and Scala certification training is designed to advance your expertise working with the Big Data Hadoop Ecosystem. You will master essential skills of the Apache Spark open source framework and the Scala programming language, including Spark Streaming, Spark SQL, Machine Learning Programming, GraphX programming and Shell Scripting Spark. This Scala and Spark certification course will give you vital skill sets and a competitive advantage for an exciting career as a Hadoop Developer.

Key Learning Objectives

- Understand the limitations of MapReduce and the role of Spark in overcoming these limitations
- Understand the fundamentals of the Scala programming language and its features
- Explain and master the process of installing Spark as a standalone cluster
- Develop expertise in using Resilient Distributed Datasets (RDD) for creating applications in Spark
- Master Structured Query Language (SQL) using SparkSQL
- Gain a thorough understanding of Spark streaming features
- Master and describe the features of Spark ML programming and GraphX programming

STEP
1
2
3
4
5
6
7
8
9
10

Course curriculum

- ✓ Lesson 1 - Introduction to Spark
- ✓ Lesson 2 - Introduction to Programming in Scala
- ✓ Lesson 3 - Using RDD for Creating Applications in Spark
- ✓ Lesson 4 - Running SQL Queries Using Spark SQL
- ✓ Lesson 5 - Spark Streaming
- ✓ Lesson 6 - Spark ML Programming
- ✓ Lesson 7 - Spark GraphX Programming

Elective Course

Scala for Data Science

This course will let flex your Scala skills for data preparation, feature engineering, creating data pipelines, and solving Big Data analytics problems. You will learn to leverage the integration of Apache Spark and Scala and how use Spark's machine learning pipelines to fit models and search for optimal hyperparameters using Scala in a Spark cluster.

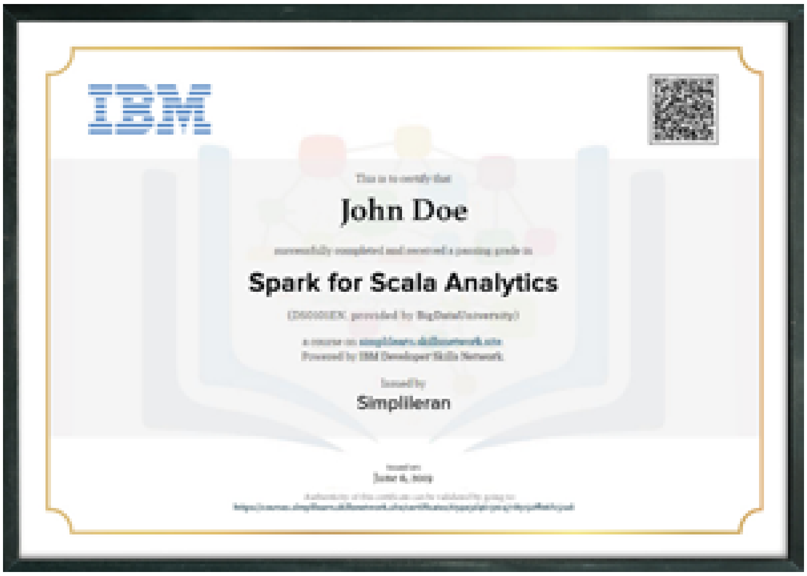
Spark for Scala Analytics

Through this course you will get an overview on the history of Apache Spark, how it evolved, how to build applications with Spark, RDDs and Data frames, Spark and its associated ecosystems. It will teach you to leverage the core RDD and DataFrame APIs to perform analytics on datasets with Scala.

Simplifying Data Pipelines with Apache Kafka

Apache Kafka is an open-source stream processing platform and a high-performance real-time messaging system that can process millions of messages per second. This Kafka training course curated by IBM will guide you through Kafka architecture, installation, interfaces, and configuration on their way to learning the advanced concepts of Big Data. It will give you hands-on experience connecting Kafka to Spark and working with Kafka Connect.

Certificates



IBM

This is to certify that

John Doe

successfully completed and received a passing grade in

Spark for Scala Analytics

(DS0101EN, provided by BigDataUniversity)

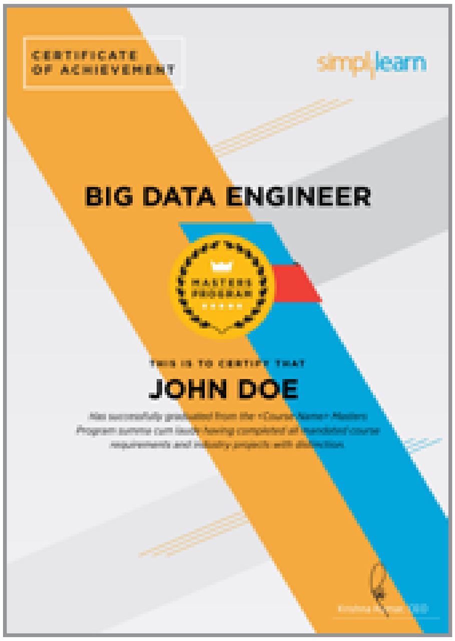
a course on simplilearn.skillsnetwork.site
Powered by IBM Developer Skills Network

Issued by

Simplileran

Issued on
June 6, 2019

Authenticity of this certificate can be validated by going to:
[illegible]



CERTIFICATE OF ACHIEVEMENT

simplilearn

BIG DATA ENGINEER

MASTERS PROGRAM

THIS IS TO CERTIFY THAT

JOHN DOE

Has successfully graduated from the <Course Name> Masters Program summa cum laude having completed all mandated course requirements and industry projects with distinction.

Krishna Kumar, CEO

Upon completion of this Master’s Program, you will receive the certificates from IBM and Simplilearn in the Big Data Engineer courses in the learning path. These certificates will testify to your skills as an expert in Data Engineering. Upon program completion, you will also receive an industry recognized Master’s Certificate from Simplilearn.

Advisory board member

Ronald Van Loon

Top 10 Big Data & Data Science Influencer, Director - Adversitement

Named by Onalytica as one of the three most influential people in Big Data, Ronald is also an author for a number of leading Big Data and Data Science websites, including Datafloq, Data Science Central, and The Guardian. He also regularly speaks at renowned events.

simplilearn

USA

Simplilearn Americas, Inc.
201 Spear Street, Suite 1100, San Francisco, CA 94105
United States
Phone No: +1-844-532-7688

INDIA

Simplilearn Solutions Pvt Ltd.
# 53/1 C, Manoj Arcade, 24th Main, Harlkunte
2nd Sector, HSR Layout
Bangalore - 560102
Call us at: 1800-212-7688

www.simplilearn.com