BIG DATA ENGINEER MASTER'S PROGRAM

In collaboration with IBM

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 certificates 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

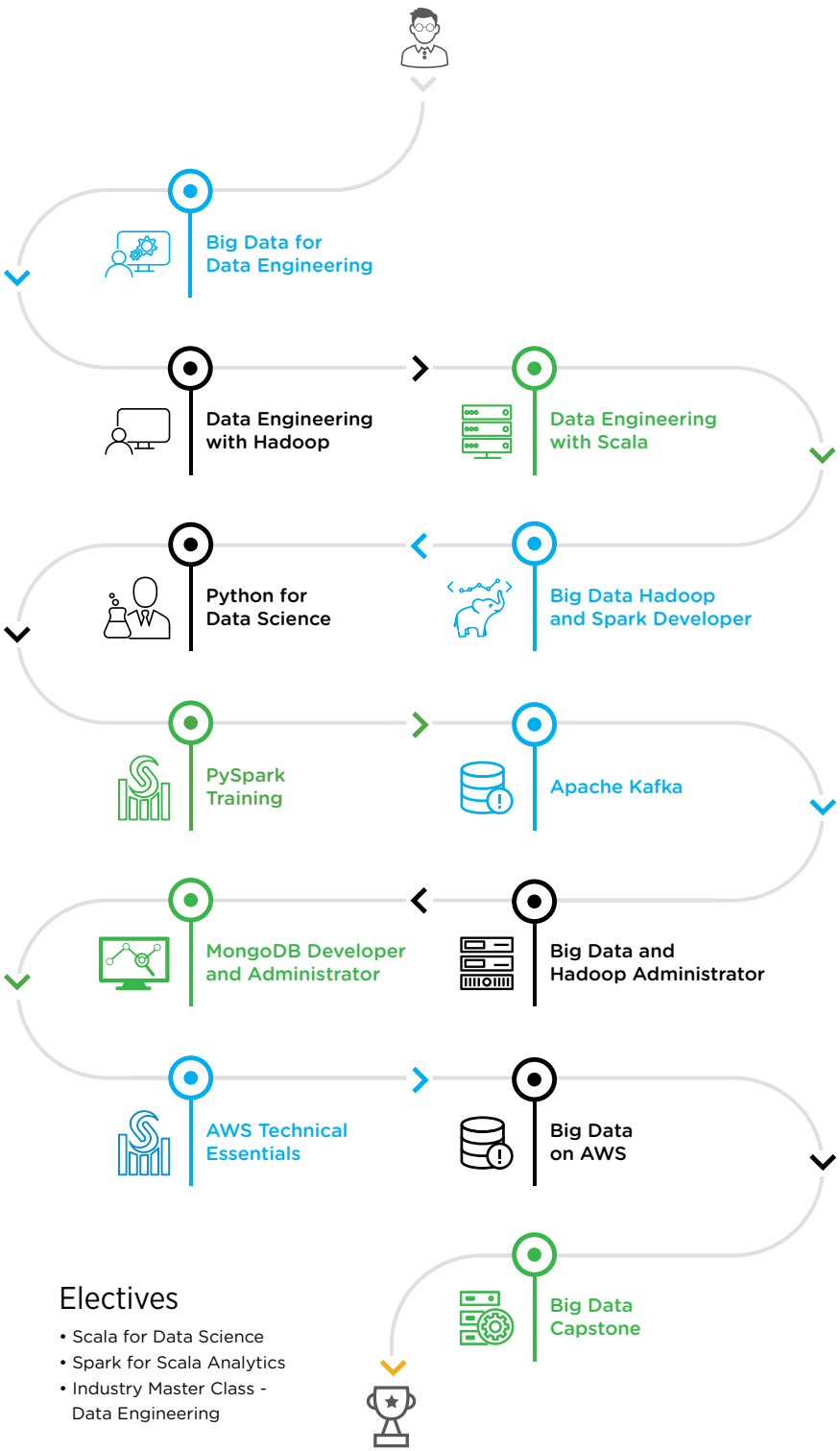
About IBM and Simplilearn collaboration

This joint partnership between Simplilearn and IBM introduces students to our integrated, Blended Learning approach, 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, 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 U.S. National Medals of Technology, five U.S. National Medals of Science, six Turing Awards, and 10 inductions into the U.S. 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 is 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 - Big Data Engineer



Electives

- Scala for Data Science
- Spark for Scala Analytics
- Industry Master Class - Data Engineering

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 such as data model creation, database interfaces, advanced architecture, Spark, Sala, 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 the 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

Understand how to use Amazon EMR for processing data using Hadoop ecosystem tools

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 extracting valuable information from them.

Key Learning Objectives

- Understand what big data is, sources of big data, and real-life examples
- Learn the key difference between big data and data science
- Master the usage of big data for operational analysis and better customer service
- Gain knowledge of the ecosystem of big data and the 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 Hadoop Distributed File System (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 (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, and Flume.

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. Upon completion of this course, carefully crafted by IBM, you will be able to write 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, including pattern matching, applying synthetic methods, and 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, and Hadoop Frameworks, including additional concepts of the big data processing life cycle. Throughout this online, instructor-led Hadoop training, you will work on real-time projects in retail, tourism, finance, and other domains. 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. Upon completion of this course, carefully crafted by IBM, 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, and conditions
- Understand the nuances of lists, sets, dictionaries, conditions and branching, objects, and classes
- Work with data in Python, including 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. This course will show you how to build and implement data-intensive applications as you dive into the world of high-performance machine learning. You'll learn how to leverage 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, understand RDDs and how to 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 01 - 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

Apache Kafka

In this Apache Kafka certification course, you will master the architecture, installation, configuration, and interfaces of Kafka open-source messaging. With this Kafka training, you will learn the basics of Apache ZooKeeper as a centralized service and develop the skills to deploy Kafka for real-time messaging. This course is part of the Big Data Hadoop Architect Master's Program and is recommended for developers and analytics professionals who wish to advance their expertise.

Key Learning Objectives

- Describe the importance of big data
- Describe the fundamental concepts of Kafka
- Describe the architecture of Kafka
- Explain how to install and configure Kafka
- Explain how to use Kafka for real-time messaging

Course Curriculum:

- Lesson 1 - Getting Started with Big Data and Apache Kafka
- Lesson 2 - Kafka Producer
- Lesson 3 - Kafka Consumer
- Lesson 4 - Kafka Operations and Performance Tuning
- Lesson 5 - Kafka Cluster Architecture and Administering Kafka
- Lesson 6 - Kafka Monitoring and Schema Registry
- Lesson 7 - Kafka Streams and Kafka Connectors
- Lesson 8 - Integration of Kafka with Storm
- Lesson 9 - Kafka Integration with Spark and Flume
- Lesson 10 - Admin Client and Securing Kafka

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 big data analytics. 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 the 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

MongoDB Developer and Administrator

Become an expert MongoDB developer and administrator by gaining an in-depth knowledge of NoSQL and master the skills of data modeling, ingestion, query, sharding, and data replication. This course includes industry-based projects in the elearning 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 the MongoDB environment
- Get hands-on experience creating and managing different types of indexes in MongoDB for query execution
- Proficiently store unstructured data in MongoDB
- Develop skill sets for processing huge amounts of data using MongoDB tools
- Gain proficiency in MongoDB configuration, backup methods, and monitoring and operational strategies
- Acquire an in-depth understanding of how to manage DB Notes, Replica set, and 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

AWS Technical Essentials

This AWS Technical Essentials course teaches you how to navigate the AWS management console; understand AWS security measures, storage, and database options; and gain expertise in web services like RDS and EBS. This course, prepared in line with the latest AWS syllabus, will help you become proficient in identifying and efficiently using AWS services.

Key Learning Objectives

- Understand the fundamental concepts of AWS platform and cloud computing
- Identify AWS concepts, terminologies, benefits, and deployment options to meet business requirements
- Identify deployment and network options in AWS

Course curriculum

- Lesson 01 - Introduction to Cloud Computing
- Lesson 02 - Introduction to AWS
- Lesson 03 - Storage and Content Delivery
- Lesson 04 - Compute Services and Networking
- Lesson 05 - AWS Managed Services and Databases
- Lesson 06 - Deployment and Management

Big Data on AWS

In this AWS Big Data certification course, you will become familiar with concepts cloud computing and its deployment models; the Amazon web services cloud platform; Kinesis Analytics; AWS big data storage, processing, analysis, visualization, and security services; EMR; AWS Lambda and Glue; machine learning algorithms; and much more.

Key Learning Objectives

- Understand how to use Amazon EMR for processing the data using Hadoop ecosystem tools
- Understand how to use Amazon Kinesis for big data processing in real-time
- Analyze and transform big data using Kinesis Streams
- Visualize data and perform queries using Amazon QuickSight

Course curriculum

- Lesson 1 - AWS in Big Data Introduction
- Lesson 2 - Collection
- Lesson 3 - Storage
- Lesson 4 - Processing I
- Lesson 5 - Processing II
- Lesson 6 - Analysis I
- Lesson 7 - Analysis II
- Lesson 8 - Visualisation
- Lesson 9 - Security

Big Data Capstone

This Big Data Capstone project will give you an opportunity to implement the skills you learned throughout this program. Through dedicated mentoring sessions, you'll learn how to solve a real-world, industry-aligned big data problem. This project is the final step in the learning path and will enable you to showcase your expertise in big data to future employers.

Elective Course

Scala for Data Science

This course will let you flex your Scala skills for data preparation, feature engineering, creating data pipelines, and solving big data analytics problems. You will learn how to leverage the integration of Apache Spark and Scala and how to 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 of the history of Apache Spark, how it evolved, how to build applications with Spark, RDDs and Data frames, the Spark ecosystem, and its associated ecosystems. You will learn how to leverage the core RDD and DataFrame APIs to perform analytics on datasets with Scala.

Certificates

Upon completion of this Master's Program, you will receive 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.

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