DATA SCIENTIST MASTERS

simplilearn

Table of Contents

About the Course

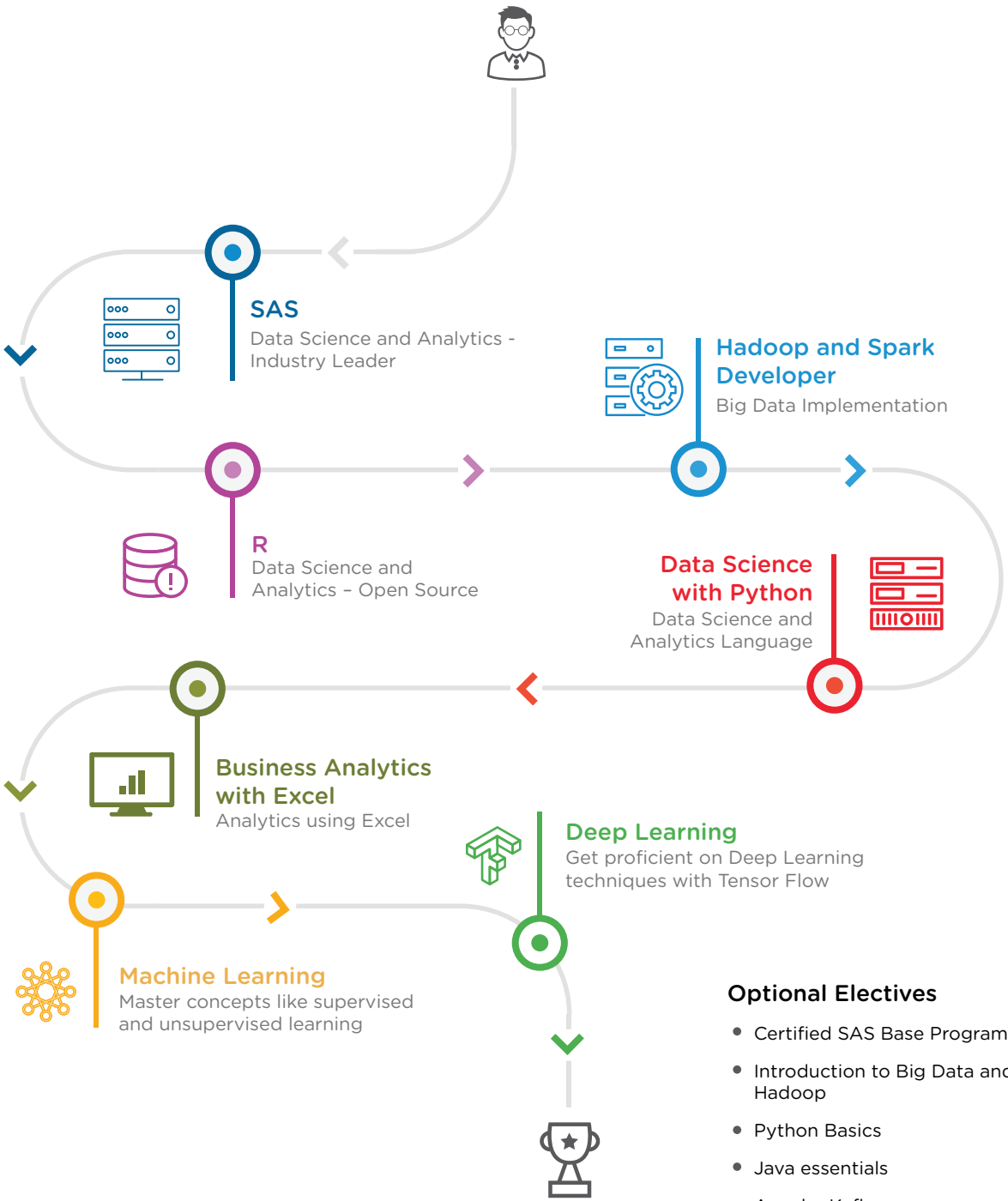
The Data Scientist Masters program has been designed to introduce you to the world of analytics and elevate your skills to ultimately become a Data Scientist. As a Data Scientist, you must be able to work with multiple data formats, have knowledge of the algorithms that can help you extract useful data, master data mining, data management and data exploration. If you are pursuing a career in Data Science, this is the program for you.

The Data Scientist Masters trains you along an industry recommended learning path to succeed in the field of Data Science

Key Features

- Industry Recommended Learning path
- Access to 200+ hours of Instructor Led training
- Hands on project execution on CloudLabs
- An industry recognized Simplilearn Masters Certificate on completion
- 17+ Real-Life projects
- GTA Technical, Project, Programming support by industry experts
- Course Advisor - Ronald van Loon

Learning Path



Optional Electives

- Certified SAS Base Programmer
- Introduction to Big Data and Hadoop
- Python Basics
- Java essentials
- Apache Kafka

DATA SCIENTIST

STEP 1 2 3 4 5 6 7

Start your Analytics journey

SAS

As a first step in the field of analytics it is important to understand various statistical concepts and learn the latest technology to apply these tools. This is why SAS is the first step in the Data Scientist Masters program. After completing this course you will be ready to work on analytics projects in some of the biggest companies in the world, as SAS is one of the most popular Data Science and analytics platforms.

Key Learning Objectives

After completing this course, you will be able to:

- Outline data science principles and how SAS can help implement them
- Explain the different methods used to combine and modify datasets
- Explain what PROC SQL is and how it's used to retrieve data from tables
- Describe how to use the macro function to manipulate the character strings and text.
- List the various statistical procedures and explore the various testing techniques.
- Understand how SAS handles missing values in your datasets using various procedures.
- Explain the ways to create a cluster and to perform cluster analysis on the dataset.
- List the various time series models of SAS.

Course Curriculum

- Analytics Overview
- SAS Introduction
- Combining/Modifying Datasets
- PROC SQL
- SAS Macros
- Basic Statistics
- Basic Statistical Procedures
- Data Exploration
- Advanced Statistical Techniques
- Working with Time series Data
- Data Optimization using SAS

STEP 1 2 3 4 5 6 7

Learn the concepts of Data Science

R

The next step to a data scientist is learning R - the upcoming and most in-demand open source technology. R is is an extremely powerful data science and analytics language which has a steep learning curve and a very vibrant community. This is why it is quickly becoming the technology of choice for organizations who are adopting the power of analytics for competitive advantage.

Key Learning Objectives

Simplilearn's Data Scientist with R programming certification has a clear focus on the vital concepts of business analytics and R programming. By the end of training, participants will be able to:

- Work on data exploration, data visualization, and predictive modeling techniques with ease.
- Gain fundamental knowledge on analytics and how it assists with decision making.
- Work with confidence using the R language.
- Understand and work on statistical concepts like linear & logistic regression, cluster analysis, and forecasting.
- Develop a structured approach to use statistical techniques and the R language.
- Perform sharp data analysis to make business decisions."

Course Curriculum

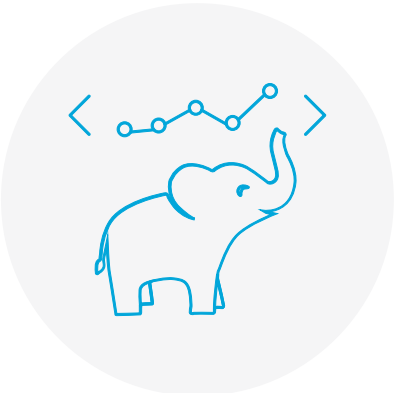
- Introduction to Analytics
- Statistical Concepts And Their Application In Business
- Basic Analytic Techniques - Using R
- Data Exploration
- Data Visualization
- Diagnostic Analytics
- Linear Regression
- Logistic Regression
- Cluster Analysis
- Time Series Analysis

STEP 1 2 3 4 5 6 7

Harness the power of Big Data & Hadoop

Big Data Hadoop and Spark Developer

The final step to a Data Scientist is the ability to work with Big Data and its components. This training is deep dive into Hadoop and all its ecosystem components including MapReduce, HDFS, Yarn, HBase, Impala, Sqoop and Flume. It also provides an introduction to Apache Spark which is a next step after Hadoop. After completing this program not only will you be ready to enter the Big Data domain but will also be able to clear the in demand Cloudera CCA175 certification.



Key Learning Objectives

- Master the concepts of the Hadoop framework and its deployment in a cluster environment
- Understand how the Hadoop ecosystem fits in with the data processing lifecycle
- Learn to write complex MapReduce programs
- Describe how to ingest data using Sqoop and Flume
- Get introduced to Apache Spark and its components
- List the best practices for data storage
- Explain how to model structured data as tables with Impala and Hive

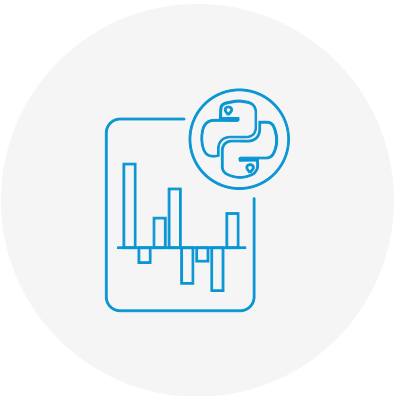
Course Curriculum

- Introduction to Bigdata and Hadoop Ecosystem
- HDFS and Hadoop Architecture
- MapReduce and Sqoop
- Basics of Impala and Hive
- Working with Impala and Hive
- Type of Data Formats
- Advance HIVE concept and Data File Partitioning
- Apache Flume and HBase
- Apache Pig
- Basics of Apache Spark, RDDs in Spark and Applications

STEP 1 2 3 **4** 5 6 7

Data Science with Python

Python is a general purpose multi-paradigm programming language for data science that has gained wide popularity because of its syntax simplicity and operability on different eco-systems.This Python course can help programmers play with data by allowing them to do anything they need- data munging, data wrangling, website scraping, web application building, data engineering and more. The Python language makes it easy for programmers to write maintainable, large scale, and robust code.



Key Learning Objectives

- Gain an in-depth understanding of data science process, data wrangling, data exploration, data visualization, hypothesis building, and testing. You will also learn the basics of statistics.
- Understand the essential concepts of Python programming like data types, tuples, lists, dicts, basic operators, and functions.
- Perform high-level mathematical, scientific and technical computing using NumPy, SciPy packages
- Perform data analysis and manipulation using data structures and tools provided in Pandas package
- Gain expertise in machine learning using the Scikit-Learn package

Course Curriculum

- Data Science Overview
- Data Analytics Overview
- Statistical Analysis and Business Applications
- Python Environment Setup and Essentials
- Mathematical Computing with Python (NumPy)
- Scientific computing with Python (Scipy)
- Data Manipulation with Pandas
- Machine Learning with Scikit-Learn
- Natural Language Processing with Scikit Learn
- Data Visualization in Python using matplotlib
- Web Scraping with BeautifulSoup
- Python integration with Hadoop MapReduce and Spark

STEP 1 2 3 4 **5** 6 7

Business Analytics with Excel

This is the basic step for all working professionals as excel is the tool of choice in most organizations. It is important for an analyst to be comfortable in excel and basic analytical concepts. These techniques help solve many of the simpler analytic problems and is an important skill for every analytics professional.

Key Learning Objectives

- Understand the meaning of business analytics and its importance in the industry
- Grasp fundamentals of excel analytics functions and conditional formatting
- Learn how to analyze with complex datasets using pivot tables and slicers
- Solve stochastic and deterministic analytical problems using tools like scenario manager, solver and goal seek
- Apply statistical tools and concepts like moving average, hypothesis testing, ANOVA and regression to data sets using Excel
- Represent your findings using charts and dashboards
- Get introduced to the latest Microsoft analytic and visualization tools i.e. Power BI

Course Curriculum

- Introduction to Business Analytics
- Formatting Conditional Formatting and Important Functions
- Analyzing Data with Pivot Tables
- Dashboarding
- Business Analytics With Excel
- Data Analysis Using Statistics
- Power BI

STEP 1 2 3 4 5 **6** 7

Machine Learning

Simplilearn's Machine Learning course will make you an expert in machine learning, a form of artificial intelligence that automates data analysis to enable computers to learn and adapt through experience to do specific tasks without explicit programming. You will master machine learning concepts and techniques including supervised and unsupervised learning, mathematical and heuristic aspects, hands-on modeling to develop algorithms and prepare you for the role of Machine Learning Engineer.

Key Learning Objectives

- Master the concepts of supervised and unsupervised learning
- Gain practical mastery over principles, algorithms, and applications of machine learning through a hands-on approach which includes working on 28 projects and one capstone project.
- Acquire thorough knowledge of the mathematical and heuristic aspects of machine learning.
- Understand the concepts and operation of support vector machines, kernel SVM, naive bayes, decision tree classifier, random forest classifier, logistic regression, K-nearest neighbors, K-means clustering and more.
- Comprehend the theoretical concepts and how they relate to the practical aspects of machine learning.
- Be able to model a wide variety of robust machine learning algorithms including deep learning, clustering, and recommendation systems

Course Curriculum

- Introduction to Artificial Intelligence and Machine Learning
- Techniques of Machine Learning
- Data Preprocessing
- Math Refresher
- Regression
- Classification
- Unsupervised learning - Clustering
- Introduction to Deep Learning

STEP 1 2 3 4 5 6 7

Deep Learning:

This Deep Learning course will transform you into an expert in deep learning techniques using TensorFlow, the open-source software library designed to conduct machine learning & deep neural network research. With our deep learning course, you'll master deep learning and TensorFlow concepts, learn to implement algorithms, build artificial neural networks and traverse layers of data abstraction to understand the power of data and prepare you for your new role as deep learning scientist.

Key Learning Objectives

- Understand the concepts of TensorFlow, its main functions, operations and the execution pipeline
- Implement deep learning algorithms, understand neural networks and traverse the layers of data abstraction which will empower you to understand data like never before
- Master and comprehend advanced topics such as convolutional neural networks, recurrent neural networks, training deep networks and high-level interfaces
- Build deep learning models in TensorFlow and interpret the results
- Understand the language and fundamental concepts of artificial neural networks
- Troubleshoot and improve deep learning models
- Build your own deep learning project
- Differentiate between machine learning, deep learning and artificial intelligence

Course Curriculum

- Introduction to TensorFlow
- Perceptrons
- Activation Functions
- Artificial Neural Networks
- Gradient Descent and Backpropagation
- Optimization and Regularization
- Intro to Convolutional Neural Networks
- Intro to Recurrent Neural Networks
- Deep Learning applications

Other Electives:

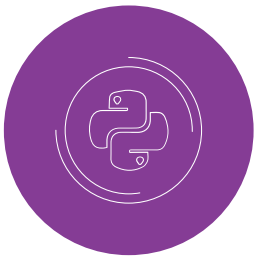
Certified SAS Base Programmer

The SAS Base Programmer course in a beginner level course for a SAS professional. This training has been designed to enable you to start your analytics career with SAS and prepare for the SAS Base Programmer certification. This SAS course explores the SAS tool and different techniques to help you access and manage data, create data structures, generate reports, and handle errors. These techniques are mandatory for a professional to start working on the next SAS assignment and forms a strong base for advanced techniques and certifications.

Introduction to Big Data and Hadoop

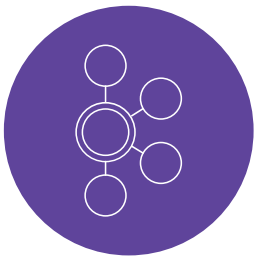
This is an ideal course for individuals who want to understand the basic concepts of Big Data and Hadoop. On completing this course, learners will be able to interpret what goes behind the processing of huge volumes of data as the industry switches over from excel-based analytics to real-time analytics.

Python Basics



This course is ideal for you to understand the basics of Python Programming Language.

Apache Kafka



This course introduces you to Apache Kafka which is an important high-performance real-time messaging system that can process millions of messages per second. It provides a distributed and partitioned messaging system and is highly fault tolerant.

Advisory Board Members:

Named by Onalytica as one of the 3 most influential people in Big Data, Ronald is an author for a number of leading Big Data & Data Science websites, including Datafloq, Data Science Central, and The Guardian. He is also a renowned speaker at industry events.

Ronald Van Loon
Big Data Expert, Director Adversitement

Sina has over 10 years of experience in Technology as a Big Data Architect at Bell Labs and as a Platinum level trainer. He is very passionate about building a Big Data education ecosystem and has been a contributor to a number of magazine and journal publications.

Sina Jamshidi
Big Data Lead at Bell Labs

Simon is a Data Scientist with 12 years of experience in Healthcare analytics. He has a master's degree in Biostatistics from the University of Western Ontario. He is passionate about teaching Data Science, and has a number of journal publications in preventive medicince and data analytics.

Simon Tavasoli
Analytics Lead at Cancer Care Ontario

Paul is a lead SAS Data Scientist at the Bank of Montreal. As an SAS Certified Predictive Modeler, SAS Statistical Business Analyst, and SAS Certified Advanced Programmer, Paul is passionate about sharing his knowledge on how Data Science can support data-driven business decisions.

Paul Sharkov
Data Scientist at BMO Financial Group, Member of SAS Canada Community

Alvaro is a Data Scientist who founded Quant Company. He has also worked as a lead Economic Analyst in the Central Bank of Guatemala. He has a master's degree in Quantitative Economics and Applied Mathematics and is actively involved in consulting and training in the Data Science space.

Alvaro Fuentes
Founder and Data Scientist at Quant Company

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
