

Integrated Program In

BIG DATA and DATA SCIENCE



simplilearn

Table of Contents



About the Course

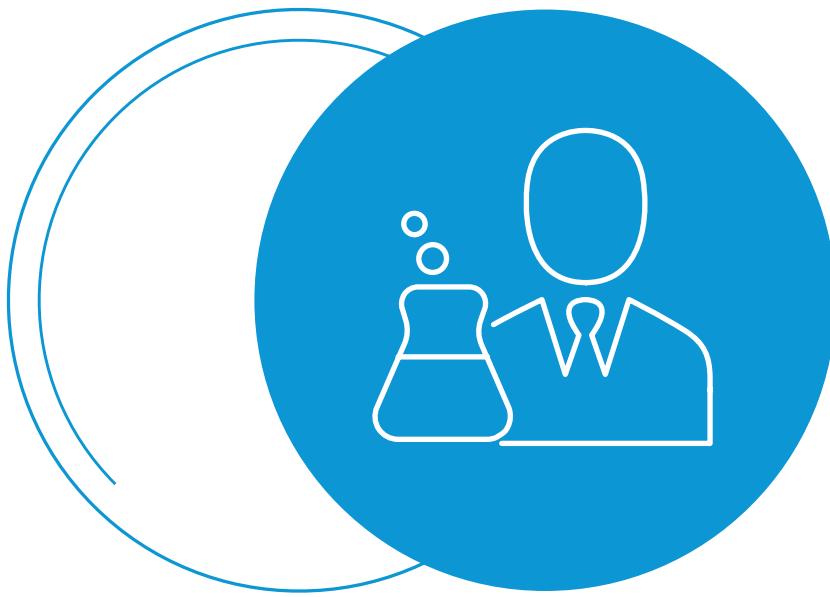


Key Features of Integrated Program in
Big Data and Data Science



Learning Path

- 06 • Get Ready for the Market
- 10 • **Step 1** : Data Science with R
- 12 • **Step 2** : Big Data Hadoop and Spark Developer
- 15 • **Step 3** : Tableau Desktop
- 17 • **Step 4** : Data Science with Python
- 19 • **Step 5** : Machine Learning
- 21 • **Electives**



About the Course

Integrated Program in Big Data and Data Science is an all-inclusive course for Big Data and data science enthusiasts. It spans across all the major technologies in big data, data science, and reporting/ visualization. This learning path is designed by some of the industry experts and big data influencers to maximize your career potential at each step. We suggest that you follow the learning path as it ensures a step-by-step transition to important concepts which prepares you for the future courses.

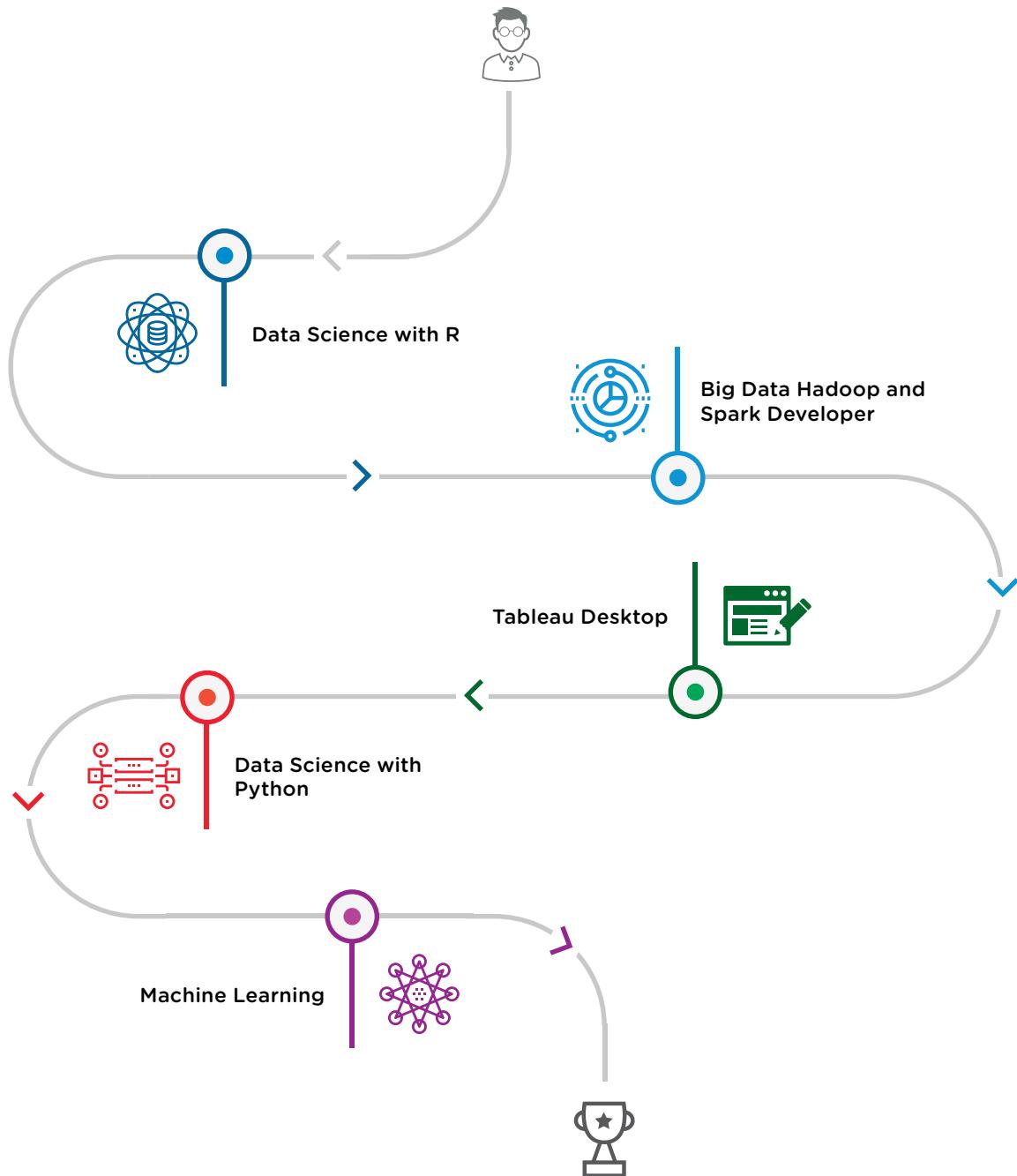
Integrated Program in Big Data and Data Science is designed to help you gain expertise in Analytics along with the latest advancements in the world of Big Data and data science.



Key Features

-  Industry-recommended learning path
-  Access to 200+ hours of instructor-led training
-  Over 250 hours of high quality e-learning
-  Hands-on project execution on CloudLabs
-  Prepares you for Cloudera CCA175 Certification and Tableau Desktop 10 Associate Certification
-  Industry recognized Simplilearn Integrated Big Data and Data Science Certificate on completion
-  15+ industry projects

Learning Path



BIG DATA AND DATA SCIENCE

Get Ready for the Market

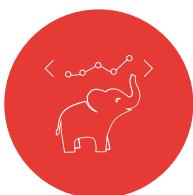
This learning path is designed for a professional taking his first steps in the world of analytics and wishes to develop skills in both big data and data science.

Key Learning Objectives



Data Science with R

This course trains you in R programming language and all the important statistical and predictive analytics concepts



Big Data Hadoop and Spark Developer

The course enables you to master the various components of Hadoop and Spark ecosystem. The course is aligned to Cloudera CCA175 certification.

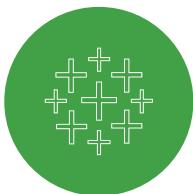
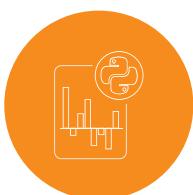


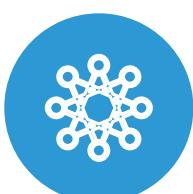
Tableau Desktop and Visualization Training

The course helps you master the various aspects of Tableau Desktop and prepares you for the Tableau Desktop Qualified Associate certification



Data Science with Python

This training introduces the various packages in Python like NumPy, SciPy, Pandas, and Scikit-learn for performing data analysis.



Machine Learning

This course helps you gain an understanding of Machine Learning applications and algorithms. It also covers deep learning and Spark Machine learning.

STEP

12345

Become an expert in the most in demand analytics programming language

Data Science with R

The Data Science with R training course has been designed to impart an in-depth knowledge of the various data analytics techniques that can be performed using R. The course is packed with real-life projects, case studies, and includes R CloudLabs for practice.



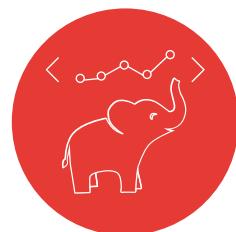
Key Learning Objectives

- ✓ Gain a foundational understanding of business analytics.
- ✓ Master the R programming and how various statements are executed
- ✓ Gain an in-depth understanding of data structure used in R and learn to import/export data in R.
- ✓ Define and use the various apply functions and DPLYR functions.
- ✓ Understand and use the various graphics in R for data visualization.
- ✓ Gain a basic understanding of the various statistical concepts.
- ✓ Understand the hypothesis testing method to drive business decisions.
- ✓ Understand regression models and classification techniques
- ✓ Learn and use the various association rules and the Apriori algorithm.
- ✓ Learn and use clustering methods including K-means, DBSCAN, and hierarchical clustering.

Be proficient in the latest advancements in Big data

Big Data Hadoop & Spark Developer

The Big Data Hadoop and Spark developer course has been designed to impart an in-depth knowledge of Big Data processing using Hadoop and Spark. The course contains with real-life projects and case studies to be executed in CloudLab and prepares you for the Cloudera CCA175 certification.



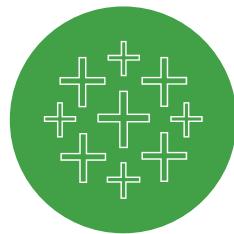
Key Learning Objectives

- ✓ Understand the architecture of HDFS and YARN, and learn how to work with them for storage and resource management
- ✓ Understand MapReduce and its characteristics
- ✓ Get an overview of Sqoop and Flume and how to ingest data
- ✓ Create databases and tables in Hive and Impala, understand HBase, and use Hive and Impala for partitioning
- ✓ Understand Flume architecture, sources, sinks and configurations
- ✓ Understand HBase, its architecture, data storage, and working
- ✓ Gain a working knowledge of Pig and its components
- ✓ Perform functional programming in Spark, understand RDDs and build spark applications
- ✓ Learn Spark SQL, and learn about creating, transforming, and querying Data frames
- ✓ Prepare for Cloudera Big Data CCA175 certification

Master Data visualization with Tableau desktop

Tableau Desktop 10

Tableau Desktop 10 training provided by Simplilearn ensures you are ready to take up a job assignment requiring Tableau Desktop expertise. The focus of the course is to help you learn Tableau Desktop 10 skills, such as visualization building, analytics, and dashboards. This course also ensures you are well prepared to clear Tableau Desktop 10 Qualified Associate exam.



Key Learning Objectives

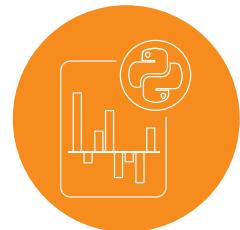
With Simplilearn's training on Tableau Desktop 10, you will be able to

- ✓ Grasp the concepts of Tableau Desktop 10 and become proficient with Tableau statistics and building interactive dashboards
- ✓ Master data connections as well as organizing and simplifying data
- ✓ Become expert in formatting, annotations, and spatial analysis
- ✓ Master Special Field Types and Tableau Generated Fields
- ✓ Understand the concepts of using charts including Pareto, waterfall, Gantt, box plots, Sparkline and perform market basket analysis
- ✓ Become expert in fundamental calculations along with automatic and custom split, ad-hoc analytics, and LOD calculations
- ✓ Master process of creating and using Parameters and gain command over mapping concepts such as custom geocoding and radial selections

Master Data Science with Python

Data Science with Python

Become an expert in data analytics, machine learning, and web scraping using Python programming. Gain an in-depth understanding of the various packages in Python like NumPy, SciPy, Pandas, and Scikit-learn for performing data analysis, implementing machine learning models, and NLP. This course is suited for both beginners and experienced professionals.



Key Learning Objectives

- ✓ Gain an in-depth understanding of data wrangling, data exploration, data visualization, hypothesis building, and testing
- ✓ Understand the essential concepts of Python programming like data types, tuples, lists, dicts, basic operators, and functions
- ✓ Perform high-level mathematical computing using NumPy package and its large library of mathematical functions
- ✓ Perform scientific and technical computing using SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO, and Weave.
- ✓ Perform data analysis and manipulation using data structures and tools provided in Pandas package
- ✓ Gain expertise in machine learning using the Scikit-Learn package
- ✓ Use matplotlib library of Python for data visualization
- ✓ Extract useful data from websites by performing web scrapping
- ✓ Integrate Python with Hadoop, Spark, and MapReduce

STEP 1 2 3 4 5

Be a Machine learning expert

Machine Learning Advanced Certification Training

This course provides advanced-level training on Machine Learning applications and algorithms. It will give you hands-on experience in multiple, highly sought-after machine learning skills in both supervised and unsupervised learning. This machine learning training ensures you can apply machine learning algorithms like regression, clustering, classification, and recommendation. The unique case study approach ensures you are working hands-on with data while you learn. You'll also receive training in deep learning and Spark Machine learning—skills which are in great demand today.



Key Learning Objectives

- ✓ Classify the types of learning including supervised and unsupervised
- ✓ Identify the various applications of machine learning algorithms
- ✓ Perform supervised learning techniques: linear and logistic regression
- ✓ Understand classification data and models
- ✓ Use unsupervised learning algorithms including deep learning, clustering, and recommendation systems
- ✓ Use machine learning with Spark

Other Electives

Data Science with SAS

The data science with SAS certification training is designed to impart an in-depth knowledge of SAS programming language, SAS tools, and various advanced analytics techniques.

[https://www.simplilearn.com/big-data-and-analytics/
data-scientist-certification-sas-excel-training](https://www.simplilearn.com/big-data-and-analytics/data-scientist-certification-sas-excel-training)



Apache Spark and Scala

With this Apache Spark certification you will master the essential skills such as Spark Streaming, Spark SQL, Machine Learning Programming, GraphX Programming, Shell Scripting Spark.

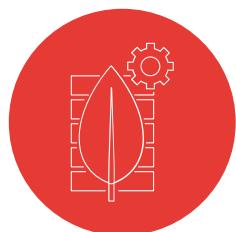
[https://www.simplilearn.com/big-data-and-analytics/
apache-spark-scala-certification-training](https://www.simplilearn.com/big-data-and-analytics/apache-spark-scala-certification-training)



MongoDB Developer and Administrator

MongoDB training helps you become job ready by mastering data modelling, ingestion, query and Sharding, Data Replication with MongoDB along with installing, updating, and maintaining MongoDB environment.

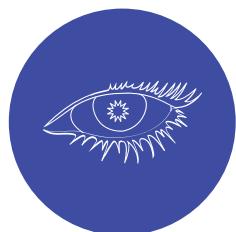
[https://www.simplilearn.com/big-data-and-analytics/
mongodb-certification-training](https://www.simplilearn.com/big-data-and-analytics/mongodb-certification-training)



Cassandra

The Apache Cassandra Training from Simplilearn provides you with in depth knowledge of Cassandra architecture, features, configuration and hadoop ecosystem around this NoSQL database

[https://www.simplilearn.com/big-data-and-analytics/
apache-cassandra-certification-training](https://www.simplilearn.com/big-data-and-analytics/apache-cassandra-certification-training)



Business Analytics with Excel

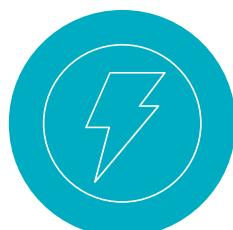
Business Analytics with Excel training has been designed to help initiate you to the world of analytics. For this we use the most commonly used analytics tool—Microsoft Excel. The training will equip you with all the concepts and hard skills required to kick-start your analytics career.



<https://www.simplilearn.com/big-data-and-analytics/business-analytics-certification-training>

Apache Storm Certification Training

Apache Storm Certification Training from Simplilearn equips you with an experience in stream processing Big Data technology of Apache Storm.



<https://www.simplilearn.com/big-data-and-analytics/apache-storm-tutorial-and-training>

Impala: An Open Source SQL Engine for Hadoop Training Course

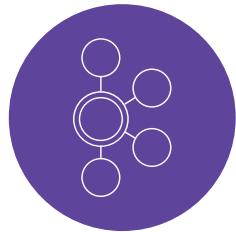
The “Impala: An Open Source SQL Engine for Hadoop” is an ideal course package for individuals who want to understand the basic concepts of Massively Parallel Processing or MPP SQL query engine that runs on Apache Hadoop. On completing this course, learners will be able to interpret the role of Impala in the Big Data Ecosystem.



<https://www.simplilearn.com/big-data-and-analytics/impala-open-source-sql-for-hadoop-training>

Apache Kafka Certification Training

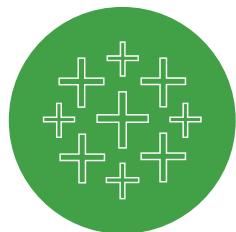
The Apache Kafka course offered by Simplilearn takes participants through the Kafka architecture, installation, interfaces, and configuration. The participants are also trained in the fundamental concepts of Big Data in this course.



<https://www.simplilearn.com/big-data-and-analytics/apache-kafka-training-tutorial>

Tableau Server 10 Qualified Associate Training

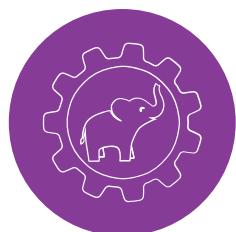
Simplilearn's Tableau Server 10 Qualified Associate course is designed to impart in-depth understanding and skills to implement, administer, and manage Tableau 10 server. This course is designed for Tableau server users and administrators.



<https://www.simplilearn.com/big-data-and-analytics/tableau-server-10-certification-training>

Big Data Hadoop Administrator

The Simplilearn Big Data and Hadoop Administrator course will prepare you for Cloudera's CCAH "CCA-500" certification and equip you with all the skills for your next Big Data admin assignment. This course covers the core Hadoop distributions—Apache Hadoop and Vendor specific distribution—CDH (Cloudera Distribution of Hadoop).



<https://www.simplilearn.com/big-data-and-analytics/big-data-and-hadoop-administrator-certification-training>

Advisory board members



Ronald Van Loon

**Top 10 Big Data & Data Science Influencer,
Director - Adversitemetn**

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



Sina Jamshidi

Big Data Lead at Bell Labs

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



Simon Tavasoli

Analytics Lead at Cancer Care, Ontario

Simon is a Data Scientist with 12 years of experience in Healthcare Analytics. He has a Masters in Biostatistics from the University of Western Ontario. Simon is passionate about teaching data science and has published several journals in preventive medicine analytics.

Advisory board members



Alvaro Fuentes

Founder and Data Scientist at Quant Company

Alvaro is a Data Scientist who founded Quant Company and has also worked as a lead Economic analyst in the Central Bank of Guatemala. He is a M.S. in Quantitative Economics and Applied Mathematics and is actively involved in consulting and training in the data science space.



Paul Sharkov

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

Paul is lead SAS Data Scientist at Bank of Montreal. As a 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.





simplilearn

USA

Simplilearn Americas, Inc.
201 Spear Street, Suite 1100, San Francisco, CA 94105
United States
Call us at: +1-844-532-7688

INDIA

Simplilearn Solutions Pvt Ltd.
53/1 C, Manoj Arcade, 24th Main, Harkunte
2nd Sector, HSR Layout
Bangalore - 560102
Call us at: 1800-102-9602