## SAS Data Scientist

# Big Data Professional

### Big Data Challenges and Analysis-Driven Data

This course provides an overview of the challenges associated with big data and analysis-driven data.

*Topics Covered*

- Reading external data files.
- Storing and processing data.
- Combining Hadoop and SAS.
- Recognizing and overcoming big data challenges.

### Exploring Data with SAS Visual Analytics

In this course, you'll learn how to use SAS Visual Analytics Explorer to explore in-memory tables from the SAS® LASR™ Analytic
Server and perform advanced data analyses.

*Topics Covered*

- Finding previously unknown relationships and spotting trends in your data.
- Visualizing data using charts, plots and tables.
- Using the autocharting function to visualize data in the best possible way.
- Using advanced graphs, such as network diagrams, Sankey diagrams and word clouds.
- Easily adding analytics to your graphs, and including descriptions of the analytics results.
- Navigating through your data using on-the-fly hierarchies.

### Statistics 1: Introduction to ANOVA, Regression and Logistic Regression

This introductory SAS/STAT® course focuses on t-tests, ANOVA and linear regression, and includes a brief introduction to logistic regression.

*Topics Covered*

- Generating descriptive statistics and exploring data with graphs.
- Performing analysis of variance and applying multiple comparison techniques.
- Performing linear regression and assessing the assumptions.
- Using regression model selection techniques to aid in the choice of predictor variables in multiple regression.
- Using diagnostic statistics to assess statistical assumptions and identify potential outliers in multiple regression.

### Preparing Data for Analysis and Reporting

In this course, you'll learn how to perform data management tasks, such as improving data quality, entity resolution and data monitoring.

*Topics Covered*

- Creating and reviewing data explorations.
- Creating and reviewing data profiles.
- Creating data jobs for data improvement.
- Establishing monitoring aspects for your data.
- Understanding the QKB components.
- Using the component editors.
- Understanding various definition types.
- Building a new data type (optional).

# SAS Data Scientist

- Using chi-square statistics to detect associations among categorical variables.
- Fitting a multiple logistic regression model.
- Scoring new data using developed models.

**Crafting Compelling (and true) Data Stories**

Storytelling is a necessary skill when talking to key stakeholders. Insights uncovered in your data can move mountains if the right people say yes.

**Introduction to SAS and Hadoop: Essentials**

This course teaches you how to use SAS programming methods to read, write and manipulate Hadoop data. You'll learn how to use Base SAS methods to read and write raw data with the DATA step, manage the Hadoop Distributed File System (HDFS) and execute MapReduce and Pig code from SAS via the HADOOP procedure. You'll also learn how to use SAS/ACCESS® Interface to Hadoop methods that allow LIBNAME access and SQL pass-through techniques to read and write Hive or Impala table structures. *Topics Covered*

- Accessing Hadoop distributions using the LIBNAME statement and the SQL pass-through facility.
- Creating and using SQL procedure pass-through queries.
- Using options and efficiency techniques for optimizing data access performance.
- Joining data using the SQL procedure and the DATA step.
- Reading and writing Hadoop files with the FILENAME statement.
- Executing and using Hadoop commands with PROC HADOOP.
- Using Base SAS procedures with Hadoop.

**DS2 Programming Essentials with Hadoop**

This course focuses on DS2, a fourth-generation SAS proprietary language for advanced data manipulation, which enables parallel processing and storage of large data with reusable methods and packages.

*Topics Covered*

- Identifying the similarities and differences between the SAS DATA step and the DS2 DATA step.
- Converting a Base SAS DATA step to DS2.
- Creating DS2 variable declarations, expressions and methods for data conversion, manipulation and conditional processing.
- Creating user-defined and predefined packages to store, share and execute DS2 methods.
- Creating and executing DS2 threads for parallel processing.
- Leveraging the SAS In-Database Code Accelerator to execute DS2 code outside of a SAS session.
- Executing DS2 code in the SAS High-Performance Analytics grid using the HPDS2 procedure.

**Big Data Analysis with Hive and Pig**

In this hands-on course, you'll use processing and analysis to find insights in structured and unstructured big data. You'll learn how to organize structural data in tabular format using Apache Hive and how to analyze the data using the Hive query language (HiveQL). You'll use the Apache Pig scripting language to perform batch processing tasks, such as extract, transform, load (ETL), data preparation and analytics.

*Topics Covered*

- Moving data into the Hadoop ecosystem.
- Using Hive to design a data warehouse in Hadoop.
- Performing data analysis using HiveQL.
- Joining data sources.

- Performing ETL.

- Organizing data in Hadoop by usage.
- Performing analysis on unstructured data using Pig.
- Joining massive data sets using Pig.
- Using user-defined functions (UDFs).
- Analyzing big data in Hadoop using Hive and Pig.

**Getting Started with SAS In-Memory Statistics**

This course focuses on accessing data on the SAS LASR Analytic Server and performing exploratory analysis and preparation. Topics include starting the server, loading data and manipulating data on the SAS LASR Analytic Server using the IMSTAT procedure. IMSTAT topics include deriving new temporary and permanent tables and columns, calculating summary statistics (e.g., mean, frequency and percentile), and creating filters and joins on in-memory data.

*Topics Covered*

- Starting up a SAS LASR Analytic Server.
- Loading tables into memory on the SAS LASR Analytic Server.
- Processing in-memory tables with PROC LASR and PROC IMSTAT.
- Accessing data more efficiently via intelligent partitioning.
- Deriving new temporary and permanent tables and variables.
- Creating filters and joins on in-memory data.
- Exporting ODS result tables for client-side graphic development.
- Producing descriptive statistics including counts, percentiles and means.
- Creating multidimensional summaries including cross-tabulations and contingency tables.
- Deriving kernel density estimates using normal functions.

# AI & ML Professional

**Module 1: Machine Learning Using SAS® Viya®**

It prepares you for the SAS Certified Specialist: Machine Learning Using SAS Viya 3.4 certification exam

Course 1: Machine Learning Using SAS® Viya®

**Module 2: Natural Language and Computer Vision**

It prepares you for the Natural Language & Computer Vision Specialist certification exam

Course 1: SAS® Visual Text Analysis in SAS® Viya®

Course 2: Deep Learning Using SAS® Software

**Module 3: Forecasting and Optimization**

It prepares you for the SAS Certified Specialist: Forecasting and Optimization Using SAS Viya 3.4 certification exam

Course 1: Forecasting Using Model Studio in SAS® Viya® 3.4

Course 2: Optimization Concepts for Data Science and Artificial Intelligence