SAS® Certified Professional: Data Curation for SAS Data Scientists

Working with SAS Data Integration Studio (20%)

Describe SAS Data Integration Studio basics
- Navigate the interface.
- Investigate global options.
- Navigate SAS folders tree.
- Investigate SAS Metadata.

Create Metadata for Source Data
- Define administrative tasks to be performed for SAS Data Integration Studio.
- Describe the New Library Wizard.
- Use Register Tables wizard to register source data.
- Register metadata for external files.

Create Metadata for Target Data and Jobs
- Describe features of the New Table wizard.
- Investigate steps for building a job.
- Discuss components of Join's Designer window.
- Create a custom transformation.
- Investigate various transformations (Extract, Summary Statistics, Join, Set Operators, Splitter)

Ensure the accuracy of data (20%)

Describe the structure of the SAS Quality Knowledge Base
- Describe the QKB component files.
- Describe various definition types.

Use DataFlux Data Management Studio
- Create and review data profiles.
- Create data jobs for data improvement.
- Apply QKB components to address data quality issues.
- Use data enrichment to create categorical data elements.

Use SAS Data Quality Server
- Configure SAS to the QKB
- Use procedures and functions

Techniques for Working with Big Data (40%)

Describe the key concepts of the Hadoop ecosystem
- Describe the Hadoop architecture
- Describe HDFS, MapReduce and YARN
- Use HUE

Query and manage Hadoop data using Hive and HiveQL
- Explain the functionality of Hive
- Use the HiveQL Data Definition Language
- Explain the Hive SerDes and storage formats
- Query Hive Table using HiveQL

Query and Manage Hadoop Data using Pig and Pig Latin
- Explain the functionality of Pig
- Describe the anatomy of a Pig script
- Use Pig Latin to manage HDFS data
- Describe Pig User Defined Functions

Access HDFS and Invoke Hadoop Applications from SAS
- Invoking Pig programs with PROC HADOOP
- Executing HDFS commands from SAS programs
- Transfer data between SAS and Hadoop
- Using data step to read and write HDFS data

Use the SAS/ACCESS SQL Pass-thru Facility
- Access Hive data using explicit SQL pass-through
- Investigate hive metadata
- Create SQL procedure pass-through queries
- Create and load hive table with SQL pass-through EXECUTE statements
- Resolve hive string data type issues in SAS

Use the SAS/ACCESS Libname Engine
- Use the SAS language to process Hive tables
- Assess and maximize performance use of the SAS/Access libname engine
- Query Hive table to create SAS reports, views, and tables
- Create Hive tables

Use DS2 Programming to manage Hadoop data

- Write DS2 programs
- Read data using DS2
- Work with variables, arrays, and ANSI SQL data types
- Use expressions and functions in DS2 programs
- Work with Methods, Packages, and Threads

Use SAS Data Loader to manage Hadoop data

- Explain the functionality of SAS Data Loader for Hadoop
- Explain the use of directives to ingest data into Hadoop
- Work with transformations to cleanse data in Hadoop
- Work with directives to blend data in Hadoop
- Create job flows using directives

Special Data Management Topics (20%)

Connect to data using SAS/ACCESS

- Describe SAS/ACCESS software
- Describe methods for accessing relational database data
- Explain performance considerations in using SAS/ACCESS methods
- Assess efficiency of SAS/ACCESS methods used during code development

Virtualize data with SAS Federation Server

- What is Data Federation, Data Virtualization, and Data Disclosure?
- Working with the SAS Federation Server
- Configuring Data Services and Connections to Disparate Data Sources
- Creating Federation SQL Views, Caches, and DS2 Queries
- Working with SAS Federation Server Security
- Accessing Data on SAS Federation Server Data

Process IoT Streams in Real Time with SAS Event Stream Processing

- Introduction to SAS Event Stream Processing
- Build SAS Event Stream Processing Models
- Use SAS Event Stream Processing Studio
- Use SAS Event Stream Processing Windows
- Use SAS Event Stream Processing Connectors and Adapters

Work with SAS Data Governance technologies

- Explain Data Governance and the Data Governance Life Cycle
- Use SAS Business Data Network to manage the business data glossary (terms, term types, and hierarchies)
- View Data Relationships in SAS Lineage