SNOWFLAKE DATA SCIENCE

Prerequisites: Basic knowledge of SQL and python is required, Foundational knowledge of databases and snowflake, A background in data science, machine learning, or statistical modeling is required.

Duration: 3 Days (8 Hrs./Day)

Course Objective: This three-day, role-specific course is intended for participants interested in developing skills and experience using Snowflake Data Cloud for data science workloads. The participant will gain exposure to the rich features of Snowflake, diverse machine learning datasets, relevant and popular open-source ML frameworks and libraries, and model deployment practices that will provide practical skills applicable to data science jobs. This course consists of lectures, demos, labs, and discussions.

Lab Requirement: Koenig DC/Linux.

Module 1 - Overview of Data Science with Snowflake

Introduction to Data Science Workload

Connecting to Snowflake

Module 2 - Snowflake Data Storage

Supported Object Types

Supported Data Types

SQL Support

Cortex ML

Cortex LLM

The Variant Data Type

Introduction to Unstructured Data

Leveraging Unstructured Data

What is Snowpark?

Module 3 - Acquire Data

Accessing External Data

Loading Data into Snowflake

Accessing Snowflake Data Worldwide with the Data Cloud

Module 4 - Prepare Data

Sampling Data

Tidying Tables

Transforming Data with Snowpark

Table Streams and Tasks

Module 5 - Perform EDA (Exploratory Data Analysis)

Tools for EDA

Univariate Regression in Snowflake

Approximation Functions

Module 6 - Perform Feature Engineering

Feature Engineering in Snowflake

Feature Engineering with Snowpark

Module 7 - Train Models

Overview of Machine Learning

Snowpark ML

Training Models with Snowpark Stored Procedures

Stored Procedures

Auto ML

Module 8 - Deploy Models

Batch Scoring

Python Worksheets

Snowflake User-Defined Functions (UDFs)

Snowpark UDFs for Model Inference

External Functions

Module 9 - Beyond Deployment: ML Ops

Improve Runtime Performance

Monitoring

ML Ops