SNOWFLAKE PYTHON DATA ENGINEER

THREE-DAY COURSE

25A31

OVERVIEW

This three-day course equips you with the expertise to integrate Python within the Snowflake AI Data Cloud. You will design and deploy high-performance data engineering solutions utilizing the Snowflake Python API and Snowpark. The course combines lectures, demos, interactive labs, and in-depth discussions to ensure a comprehensive learning experience.

ACQUIRED SKILLS

- Explain the distinctive features of Snowflake's platform and its integration with Python.
- Configure and establish secure connections to Snowflake using the Snowpark Session object.
- Design, code, and deploy custom Python functions within Snowflake as User Defined Functions (UDFs).
- Create and encapsulate reusable logic using Stored Procedures.
- Organize and manage automated workflows with Snowflake tasks and Directed Acyclic Graphs (DAGs).
- Automate recurring data tasks using Snowflake's task scheduling capabilities.
- Monitor and debug data processes while implementing observability techniques in Snowflake and Python environments.
- Leverage Anaconda integration in Snowflake to enhance data solutions with specialized Python libraries.

WHO SHOULD ATTEND

- Data Engineers
- Data Scientists
- Data Application Developers
- Database Architects
- Database Administrators
- Data Analysts with programming experience

PREREQUISITES

- Basic Python coding proficiency.
- Familiarity with basic SQL.

DELIVERY FORMAT

Instructor-led Public or Private classes are available.

TOPICS COVERED

Snowflake AI Data Cloud

- Using Snowsight
- Snowflake Structure

Snowflake Python API

- Python API Concepts
- Core Classes and Operations

Role-based Access Control (RBAC) Overview

Data Protection Features

- Cloning
- Time Travel

Metadata and Caching in Snowflake

- Metadata
- Query Result Cache
- Data Cache

Introduction to the Data Engineering Workflow

Supporting Platform Features

- Snowpark
- Snowflake Connector for Python
- Drivers, Clients, and Connectors Overview
- Snowflake Notebook API

Ingestion

- Data Loading Objects
- Transformations and Copy Options
- Bulk vs. Continuous Data Loading Approaches
- Semi-structured Data

- Snowpipe
- Snowflake Data Loading Best Practices
- Loading Semi-structured Data

Transformation

- Dynamic Tables
- Creating and Managing Streams
- UDFs and Stored Procedures
- External Network Access
- Transformations with Unstructured Data

Orchestration

- Creating Tasks
- Creating a DAG

Delivery

- Streamlit
- Data Sharing

Management and Observability

- Observability on Snowflake
- Outbound Notifications
- Snowflake Alerts
- Data Pipeline Logging