



SNOWFLAKE DATA ENGINEER

THREE-DAY COURSE

24I19

OVERVIEW

This three-day role-specific course covers key concepts, features, considerations, and Snowflake-recommended best practices through the lens of the data engineering workflow. It is intended for participants who will be accessing, developing, and querying datasets for analytic tasks and building data pipelines in Snowflake. This course consists of core data engineering concepts delivered through lectures, demos, labs, and discussions.

ACQUIRED SKILLS

- Describe the data engineering workflow and how the Snowflake AI Data Cloud features support the various components of the workflow.
- Access Snowflake through the Snowsight UI and by using application methods.
- Load and unload data sets.
- Configure Snowflake features to cover a range of data ingestion and processing latencies.
- Develop applications for Snowflake, including comprehensive ANSI standard SQL support.
- Employ performance and cost optimization techniques.
- Use Snowflake’s capabilities to work effectively with structured, semi-structured, and unstructured data in Snowflake.
- Tune queries and improve performance using advanced techniques such as data clustering and materialized views.
- Employ Snowflake SQL extensibility features such as user-defined functions and stored procedures.

WHO SHOULD ATTEND

- Data Analysts
- Data Engineers
- Data Scientists
- Database Architects
- Database Administrators
- Data Application Developers

PREREQUISITES

- A background in data engineering is required.
- Completion of “Snowflake Foundations” one-day course or equivalent Snowflake knowledge.

DELIVERY FORMAT

Instructor-led Public or Private classes are available.

TOPICS COVERED

Snowflake Data Cloud

Introduction to the Data Engineering Workflow

Supporting Platform Features

- Authentication Methods
- Drivers, Clients, and Connectors Overview
- Snowflake Connector for Python
- SnowSQL
- Role-based Access Control (RBAC) Overview
- Introduction to Data Governance

Data Storage

- Semi-structured Data
- Query Semi-structured Data
- Query Tags
- Data Lake
- Apache Iceberg Tables
- External Tables

Powering Data With Snowflake LLMs

- Document AI
- Cortex LLM Functions Overview
- Cortex LLM Functions Specialized Functions
- Cortex LLM Functions Complete
- Cost Monitoring

Ingestion Layer

- Bulk vs. Continuous Data Loading Approaches
- Snowpipe
- Snowpipe Streaming
- Snowflake Connector for Kafka
- Snowflake Connector for Kafka With Snowpipe Streaming
- Snowflake Data Loading Best Practices

- Loading Semi-structured Data
- Schema Detection
- Working With Unstructured Data
- Creating and Managing Streams
- Streams on Views

Orchestration

- Creating and Managing Tasks
- Using Streams and Tasks Together

Transformation

- Dynamic Tables
- Extensibility Overview
- Snowflake Scripting
- UDFs and UDTFs
- Extend Snowflake With Java and Python
- External Functions
- External Network Access
- Introduction to Snowpark
- Transformations With Unstructured Data

Performance Optimization

- Natural Clustering
- Explicit Clustering
- Automatic Clustering Service
- Search Optimization Service Introduction
- SQL Performance Tips
- Performance Bottleneck Scenarios

Modeled Layer

- Materialized Views
- Unloading Semi-structured Data
- Data Sharing
- Secure Views

Management and Observability

- Observability on Snowflake
- Outbound Notifications
- Snowflake Alerts
- Data Metric Functions
- System DMF
- Custom DMF
- Observability Within Snowsight
- Cost Controls
- Resource Monitors
- Working With JupyterLabs