SNOWFLAKE ADVANCED

THREE-DAY COURSE

23J16

OVERVIEW

This three-day course covers advanced Snowflake features related to data movement, performance, authentication, design considerations, and recommended best practices in the Snowflake Data Cloud. This advanced course consists of lectures, demos, and labs.

ACQUIRED SKILLS

- Evaluate Snowflake's advanced architectural concepts.
- Design a bulk loading and load troubleshooting strategy.
- Employ strategies and Snowflake-recommended best practices for continuous data processing.
- Develop a methodology for performance tuning for your Snowflake Data Cloud.
- Protect data access at the table, column, and row levels.
- Use data sharing and database replication.

WHO SHOULD ATTEND

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

PREREQUISITES

Completion of "Snowflake Fundamentals" or equivalent Snowflake knowledge.

DELIVERY FORMAT

Instructor-led Public or Private classes are available.

TOPICS COVERED

Architecture Highlights

- Architectural Layers
- Caching

Connecting to Snowflake

- What is Integration?
- Security Integration
- Key Pair Authentication

Extensibility and Programmability

- External Functions
- API Integrations
- Tools and Utilities
- Developer Framework
- Turn Python Scripts into Web Apps (Streamlit)

Working with Semi-Structured Data

- Structured vs. Semi-structured
- Loading and Transforming Semi-structured Data
- Schema Detection
- Schema Evolution

Working with Unstructured Data

- Overview
- Concepts
- Workflow

Working with Geospatial

- Geospatial Overview
- Geography
- Geometry

Data Loading and Troubleshooting

- Data Ingestion Options
- Fixing Load Failures
- Continuous Loading with Snowpipe
- Notification Integration
- Snowpipe Streaming Overview

- Snowpipe Streaming Detail

Transformation Pipelines

- Traditional Micro-batch Pipeline
- Alternative Continuous Data Pipeline
- Stream Details
- How Streams Work
- Common Use Cases
- Dynamic Tables

Working with Data Lakes

- Data Lake Deployments
- Querying External Data Lakes
- Creating and Querying External Tables
- Partitioning External Tables

Data Clustering

- What is Data Clustering?
- Micro-partition Pruning (Elimination)
- Evaluating Clustering
- Implement and Test Cluster Keys

Performance Optimization

- Search Optimization
- Materialized Views
- Materialized View Use Cases
- When to Use Materialized Views
- Query Acceleration Services (QAS)
- Join Elimination Using Constraint Support

Data Sharing

- Data Access Options
- Direct Data Sharing Workflow