DATA ENGINEER

OVERVIEW

This role specific course covers the Snowflake key concepts, features, considerations, and best practices intended for key stakeholders who will be accessing, developing, querying datasets for analytic tasks and and building data pipeline in Snowflake. These stakeholders often the role of database application developer and data engineer. This course will consist of lectures, demos, labs, and discussions.

ACQUIRED SKILLS

> Overview of Snowflake key features and architecture

> Performance and cost optimization techniques using caching and high performing functions

> Learn different UI and application methods of accessing Snowflake

> Use the capabilities and best practices for working with semi-structured data in Snowflake

> Load, unload data sets and best practices

> Tune queries and performance using advanced techniques such as data clustering and materialized views

> Develop application for Snowflake including using comprehensive ANSI standard SQL support

> Leveraging Snowflake SQL extensibility features such as time travel capabilities, user-defined functions and stored procedures

WHO SHOULD ATTEND

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

PREREQUISITES

Basic knowledge of SQL is helpful.

DELIVERY FORMAT

Instructor-led class

SCHEDULE

Snowflake Architecture and Overview

- Snowflake Technical Overview
- Review three-tiered architecture
- Cloud Services
- Virtual Warehouse
- Storage

Interfaces and Connectivity

- Using Snowflake Web Console, Worksheet
- Using Snowflake command line, SnowSQL
- Overview of connectors and ecosystem for Snowflake
- A deeper look at specific programming interface such as Python and Spark

Developing for Snowflake

- Overview of programmatic interfaces for Snowflake
- Application use case scenarios
- A deeper look at a specific programming interface such as Python and Spark

Loading New Data Sets

- Ingesting new data into Snowflake tables
- Working with various SQL data types
- Discuss streaming data
- Ingestion best practices and anti-patterns

Querying Data and Best Practices

- Review best practices of writing effective queries
- Walkthrough filtering data examples and best practices
- Walkthrough grouping and sorting data and performance considerations
- Usage and pitfalls of joining data

SCHEDULE

Query Caching Performance Features

- Result set cache
- Metadata cache
- Query data cache
- Best Practices of using caching for performance and cost optimization

Performing Data Analytic Tasks

- Using Snowflake's high performing approximation and estimation functions
- User-defined functions
- Time Travel queries

Performance monitoring and management of query and ETL workloads

- Query profiling
- Virtual warehouse (compute resource) management
- Tuning workloads
- Monitoring functions
- Optimization of storage & compute
- Optimizing cost

Using Data Clustering Optimization for Advanced Query Performance Tuning

- How to identify appropriate use cases
- Designing clustering keys
- Auto-clustering service
- DML considerations
- Materialized View use cases
- Materialized View features

Test, QA, and Production and Agile Development

- Time Travel queries in Snowflake
- Cloning data and environment in Snowflake

SCHEDULE

Working with Semi-Structured Data

- Data source formats
- Support of native data types
- SQL Operations (grouping, sorting and more)
- Built-in functions for traversing, flattening, and nesting of semi-structured data

Sharing Data in Snowflake

- Data Sharing Overview
- Secured View
- Secured UDF