



VECTOR E

SNOWFLAKE DATA ENGINEERING TRAINING

SNOWFLAKE, DBT & ADF CURRICULUM

Getting Started with Snowflake Topics

- Snowflake CLI installation & Configuration
- Different Roles in Snowflake Explore - Databases, Schemas and Tables
- Snowflake Architecture
- Snowflake Virtual warehouses overview

Using Snowflake Topics

- Web Interface
- Virtual Warehouses
- Databases, Tables & Views
- Queries
- Date & Time Data
- Semi-structured Data

Micro-partitioning and Clustering in Snowflake

- Micro partitions in Snowflake
- Advantages of Micro-Partitioning
- Structure of Micro-Partition
- Query Processing in Snowflake
- Clustering Introduction

- How clustering works & It's advantages
- how to Choose the right keys for clustering
- Performance tuning

Table types in Snowflake:

- Permanent table and fail-safe
- Transient table
- Temporary table
- Transient database in snowflake

Snowflake Continuous Data Pipelines:

- Snowflake Continuous stream data integrations, Streams,
- Tasks
- Snow pipe

Sharing Data in Snowflake:

- Introduction to Data Sharing
- Data Providers
- Data Consumers
- Working with Shares

Snowflake Task Feature

- Tasks Intro
- Demo SF
- task Dependency
- Dependency demo
- Understanding Snowflake Billing:
- Monitoring Credit and Storage Usage
- Resource Monitors
- User Management

Snowflake Concepts

- SQL in Snowflake
- Loading Data – Structured and Semi Structured
- Materialized Views
- Fail Safe and Time Travel
- Cloud Computing
- Access Control
- Zero Copy Cloning
- Swapping
- Snowflake Roles
- Window Functions
- Stored Procedures
- User Defined Functions
- SnowSQL
- Query Profile

Introduction to dbt

- dbt in modern data stack
- Snowflake with dbt – Accelerating Data Teams
- dbt Cloud vs dbt Core
- Setting up dbt Cloud
- dbt Cloud IDE Overview
- Setting up dbt Core
- dbt supported data platforms
- dbt supported version control
- Initialize dbt project
- Materializations

Models

- Models in dbt
- Model folders/layers
- Building first model
- dbt run
- config block
- Building Modularly
- ref functions
- Naming Convention of Models

Sources

- What are sources
- Configure sources
- Select from sources
- dbt source freshness

Seeds and Incremental Materialization

- Creation of seed data
- run the seed
- creation of incremental model
- run the incremental model
- change seed data
- test incremental model

dbt tests

- Tests & Significance
- Understanding datasets in data model
- Generic tests - unique, not_null, accepted_values, relationships
- Singular tests
- Testing sources
- Custom generic tests
- Implementing testing steps in workflow
- dbt build
- Data validations

dbt documentation

- Documentation significance
- Documentation in dbt

- Updating dbt docs
- doc blocks in mark down file
- Implementing source, table, and column descriptions in .yml files
- generating a documentation site using dbt docs
- model and data lineage on the DAG

Deployment & Environments

- deployment vs development environments
- Setting up environments
- Understanding environment's connections
- Creating jobs and configuring job settings
- Job Triggers - Schedule, webhooks, API

Git Version Control

- About Git & Version control
- Git branches & functionalities
- Commits & pull requests
- Merging code

dbt pipelines

- staging models
- transform models
- understanding errors
- dbt plugins
- using dbt packages

- Query Tag
- Creation of Snapshots

Additional concepts

- Jinja and macros
- Working with variables

Azure Data Factory

1. Introduction to Azure Data Factory

- Overview of Azure Data Factory (ADF)
- Key features and benefits
- Use cases and scenarios
- Hands-on Lab: Setting up your Azure Data Factory environment

2. Different Kinds of Integration Runtimes

- Definition and purpose of Integration Runtimes (IR)
- Types: Azure, Self-hosted, and Azure-SSIS Integration Runtime
- When and how to use each type
- Hands-on Lab: Setting up different types of Integration Runtimes

3. Linked Services

- What are Linked Services?
- How to create and manage Linked Services
- Examples of commonly used Linked Services
- Best practices for securing Linked Services
- Hands-on Lab: Creating and managing Linked Services

4. Data Sets

- Definition and role of Data Sets in ADF
- How to create and configure Data Sets
- Examples and best practices
- Hands-on Lab: Creating and managing Data Sets

5. Pipelines

- Introduction to Pipelines
- Components of a Pipeline
- How to create and manage Pipelines
- Best practices for designing efficient Pipelines
- Hands-on Lab: Building your first Pipeline

6. Parameters vs Variables

- Differences between Parameters and Variables
- How and when to use Parameters and Variables
- Practical examples
- Hands-on Lab: Using Parameters and Variables in a Pipeline

7. Copy Data

- Overview of the Copy Data tool
- Step-by-step guide to copying data
- Common scenarios and troubleshooting
- Hands-on Lab: Copying data between different data stores

8. Data Flow Activity

- Introduction to Data Flow in ADF
- Building and managing Data Flows
- Examples of Data Flow transformations
- Best practices for optimizing Data Flows
- Hands-on Lab: Creating a Data Flow with transformations

9. Configure Azure Integration Runtime

- Setting up Azure Integration Runtime
- Configuration options and best practices
- Common issues and resolutions
- Hands-on Lab: Configuring Azure Integration Runtime

10. Configure SSIS Integration Runtime

- Overview of SSIS IR
- Steps to configure and manage SSIS IR
- Migration tips and tricks
- Hands-on Lab: Configuring SSIS Integration Runtime

11. System Assigned and User Assigned Managed Identities

- Explanation of Managed Identities
- Differences between System Assigned and User Assigned Managed Identities
- Implementation and use cases
- Hands-on Lab: Implementing Managed Identities in ADF

12. Lookup and Metadata Activity

- Introduction to Lookup and Metadata Activity
- How to configure and use these activities
- Practical examples
- Hands-on Lab: Using Lookup and Metadata Activity in Pipelines

13. For Each Activity

- Understanding For Each Activity
- Configuration and use cases
- Best practices and examples
- Hands-on Lab: Implementing For Each Activity in a Pipeline

14. Fail Over, Validate

- Strategies for Failover in ADF
- Validation mechanisms and best practices
- Examples and common pitfalls
- Hands-on Lab: Implementing failover and validation in Pipelines

15. IF, Switch, Until

- Conditional activities in ADF: IF, Switch, Until
- How to configure and use each activity
- Practical scenarios and examples
- Hands-on Lab: Using IF, Switch, and Until activities in Pipelines

16. Set and Append Variable

- Using Set and Append Variable activities
- Configuration and best practices
- Examples and use cases
- Hands-on Lab: Setting and appending variables in a Pipeline

17. Stored Procedure

- Executing Stored Procedures in ADF
- Configuration and usage
- Best practices and troubleshooting
- Hands-on Lab: Executing a Stored Procedure in a Pipeline

18. Different Kinds of Triggers

- Overview of Triggers in ADF
- Types: Schedule, Tumbling Window, Event-based
- How to create and manage Triggers
- Best practices for trigger management
- Hands-on Lab: Creating and managing different types of Triggers

19. Monitoring and Managing Pipelines

- Tools for monitoring Pipelines
- Debugging and troubleshooting techniques
- Hands-on Lab: Monitoring and managing Pipeline runs

20. Advanced Topics

- Incremental data loading
- Handling schema changes
- Integration with other Azure services (e.g., Azure Synapse, Databricks)
- Hands-on Lab: Implementing an end-to-end data integration solution

21. Power Query Activity

- Introduction to Power Query in ADF
- Building and managing Power Query transformations
- Practical examples and use cases
- Best practices for using Power Query
- Hands-on Lab: Implementing Power Query transformations in a Pipeline

22. Real-world Case Studies

- Case study 1: Data migration from on-premises to cloud
- Case study 2: Building a data warehouse with ADF
- Case study 3: Integrating ADF with other Azure services

23. Best Practices

- Design best practices for ADF
- Security best practices
- Performance optimization tips
- Troubleshooting common issues

24. Additional Resources

- Documentation and tutorials
- Blogs and articles
- Community forums and support

25. Q&A Sessions

- Regular Q&A sessions to address participant questions
- Dedicated time for troubleshooting and discussion

26. Certification Preparation

- Overview of relevant Azure certifications
- Study tips and resources
- Practice exams and sample questions

This comprehensive curriculum covers all essential aspects of Azure Data Factory, including advanced topics and hands-on practice.