📓 Snowflake + AI Engineering Course (Complete Curriculum)

🎯 Course Objective

This course is designed to make you an **end-to-end Snowflake Data & AI Engineer**, covering:

- Data Warehousing
- Snowflake Development & Administration
- Performance & Security
- Real-time Pipelines
- AI (Cortex, RAG, Agents)

□ Module 1: Data Warehousing Fundamentals

Topics:

- OLTP vs OLAP
- Data Warehouse Architecture
- Fact & Dimension Tables
- Star Schema vs Snowflake Schema
- SCD Types (Type 1, Type 2, Type 3)
- Kimball vs Inmon Approach

☁ Module 2: Cloud Fundamentals

Topics:

- Cloud Concepts (IaaS, PaaS, SaaS)
- Storage vs Compute Separation
- Object Storage (AWS S3, Azure Blob)
- IAM Basics

❄️ Module 3: Snowflake Architecture

Topics:

- Snowflake Layers:
  - Storage Layer
  - Compute Layer (Virtual Warehouses)
  - Cloud Services Layer
- Query Execution Flow (Input → Optimization → Execution → Result)
- Multi-cluster Warehouses

🛠️ Module 4: Snowflake Core Development

Topics:

- Databases, Schemas, Tables
- Table Types:
  - Permanent
  - Transient
  - Temporary
- File Formats
- Stages:
  - Internal Stage
  - External Stage
- Data Loading:
  - COPY INTO
- Data Unloading

🔄 Module 5: Data Loading & Integration

Topics:

- Bulk Loading
- Snowpipe (Auto ingestion)
- Event-based ingestion
- Loading from:
  - AWS S3
  - Azure Blob

    - GCS
    - On-prem systems

□ Module 6: SQL & Data Transformation

Topics:

- Joins (Inner, Left, Right, Full)
- Window Functions
- CTEs & Subqueries
- Aggregations
- Semi-structured Data:
    - JSON handling
    - VARIANT data type
    - FLATTEN function

⚡ Module 7: Snowflake Advanced Features

Topics:

- Time Travel
- Zero Copy Cloning
- Fail-safe
- Data Retention Policies

🚀 Module 8: Performance Optimization

Topics:

- Micro-partitions
- Clustering & Clustering Keys
- Query Optimization Techniques
- Caching:
    - Result Cache
    - Warehouse Cache
- Query Profile Analysis

🔐 Module 9: Security & Governance

Topics:

- RBAC (Role-Based Access Control)
- Role Hierarchy
- Row-Level Security
- Column-Level Security
- Dynamic Data Masking
- Secure Views
- Network Policies
- Data Governance Concepts

👩‍💼 Module 10: Snowflake Administration

Topics:

- Warehouse Sizing & Scaling
- Resource Monitors (Cost Control)
- User & Role Management
- Query Monitoring
- Account Usage & Billing Analysis

🖇️ Module 11: Data Sharing & Collaboration

Topics:

- Secure Data Sharing
- Cross-region Sharing
- Data Marketplace

🔁 Module 12: Streams & Tasks (Real-time Pipelines)

Topics:

- Change Data Capture (CDC)
- Streams Internals
- Tasks Scheduling
- Building Pipelines using Streams + Tasks

□ Module 13: Snowflake Ecosystem

Topics:

- Snowpark (Python, Scala, Java)
- User Defined Functions (UDFs)
- Stored Procedures (SQL + Snowpark)
- External Functions

🤖 Module 14: AI with Snowflake

Topics:

- Snowflake Cortex AI Overview
- LLM Functions:
  - COMPLETE
  - SUMMARIZE
  - CLASSIFY
- Embeddings
- Vector Search
- Prompt Engineering in Snowflake

💬 Module 15: AI Chatbot & Assistant

Topics:

- Chatbot Architecture
- Context Management
- Memory Handling
- Integration with Snowflake

🔍 Module 16: RAG (Retrieval-Augmented Generation)

Flow:

Input → Embedding → Vector Search → LLM → Response

Topics:

- RAG Architecture
- Chunking Strategies
- Embeddings Pipeline
- Vector Databases
- Frameworks like LangChain

🤖 Module 17: AI Agents

Topics:

- Agent Architecture
- Tool Calling
- Multi-agent Systems
- Autonomous Workflows

🐍 Module 18: Python & Snowpark

Topics:

- Snowpark DataFrames
- Transformations
- Writing back to Snowflake
- Integration with AI Pipelines

🏗 Module 19: End-to-End Projects

🔷 Project 1: Batch Pipeline

Flow:
S3 → Snowflake → Transform → BI Tool

Includes:

- Data Loading
- Transformations
- Reporting

🔷 Project 2: Real-time Pipeline

Flow:
Snowpipe → Streams → Tasks → Final Tables

Includes:

- Auto ingestion
- CDC pipeline

🔷 Project 3: AI Chatbot with RAG

Flow:
User Query → Vector Search → LLM → Response

Includes:

- Cortex AI

- Snowpark
- RAG implementation

📊 Module 20: Interview & Real-world Scenarios

Topics:

- Architecture Design Questions
- Performance tuning scenarios
- Debugging queries
- Cost optimization strategies