D4DATA.in

FULL STACK SNOWFLAKE DATA ENGINEERING

2.5-Month Practical Program | 1.5 Hours per Class
Snowflake + dbt at the Core • SQL • Azure • Python • Databricks • GitHub

Program Objective

This program is designed to take a learner from core data engineering foundations to building production-style Snowflake solutions. The emphasis is on hands-on practice, architecture understanding, troubleshooting, and two end-to-end projects rather than only theory.

Duration **2.5 Months**	**Class Duration** **1.5 Hours**	**Projects** **2 End-to-End**
Core Focus **Snowflake + dbt**	**Learning Style** **Hands-on**	**Delivery** **Live + LMS**

Learning Journey

01	SQL Strong querying foundation
02	Azure Cloud fundamentals for data engineering
03	Snowflake + dbt Core platform, ELT, automation & modelling
04	Python Practical Python for data workflows
05	Databricks Spark-based engineering & Delta Lake
06	GitHub Version control & team workflow

Outcome: By the end of the program, candidates should be able to explain, build, test, version and troubleshoot a modern data pipeline using Snowflake-centered architecture.

Detailed Curriculum

MODULE 1 • SQL FOR DATA ENGINEERING

Goal: Build the SQL foundation required for Snowflake, analytics and data engineering interviews.

- Database basics: tables, rows, columns, keys, schemas and relational concepts
- DDL, DML, DQL and TCL fundamentals; CREATE, ALTER, DROP, INSERT, UPDATE, DELETE and SELECT
- Filtering, sorting, DISTINCT, CASE, NULL handling and built-in functions
- Aggregations: GROUP BY, HAVING, COUNT, SUM, AVG, MIN and MAX
- Joins: inner, left, right, full, cross and self joins; practical join scenarios
- Set operators: UNION, UNION ALL, INTERSECT and EXCEPT/MINUS concepts
- Subqueries, correlated subqueries, CTEs and reusable query patterns
- Window functions: ROW_NUMBER, RANK, DENSE_RANK, LEAD, LAG and running calculations
- Date, string and conditional transformations used in ETL
- Interview-style SQL problems: duplicates, nth highest salary, top-N, gaps, latest records and reconciliation
- Query readability, debugging and basic performance thinking

MODULE 2 • AZURE CLOUD FUNDAMENTALS

Goal: Understand how cloud services fit into a modern data platform.

- How applications run: compute, storage, networking and databases
- On-premises vs cloud; scalability, elasticity, availability and pay-as-you-go
- IaaS, PaaS and SaaS with practical examples
- Azure subscriptions, resource groups, regions and basic identity concepts
- Azure Storage and ADLS Gen2: containers, folders and data-lake concepts
- Azure Data Factory fundamentals: linked services, datasets, pipelines, activities and triggers
- Integration Runtime concepts and common source-to-cloud movement patterns
- Azure Key Vault / secrets concepts and secure connection thinking
- High-level Azure data engineering architecture and how Azure can connect with Snowflake and Databricks

MODULE 3 • SNOWFLAKE + DBT — CORE MODULE

This is the primary focus of the program.

- Snowflake architecture: cloud services, query processing and database storage layers
- Accounts, databases, schemas, tables, views, stages, file formats and virtual warehouses
- Roles, users, privileges and RBAC fundamentals
- Loading data: internal/external stages, COPY INTO and validation/error handling
- Snowpipe and event-driven / continuous ingestion concepts
- Micro-partitions, pruning, caching, clustering and performance fundamentals
- Warehouse sizing, auto-suspend/resume, scaling and cost-conscious design
- Semi-structured data: VARIANT, JSON, FLATTEN and practical transformations
- Streams and Tasks for change tracking and orchestration
- Dynamic Tables and when to use them
- Time Travel, Fail-safe concepts, zero-copy cloning and data recovery workflows
- Secure views, masking policies, row access concepts and governance fundamentals
- Sharing data securely and Snowflake data-sharing concepts
- Query Profile, Query History and systematic troubleshooting of slow/expensive SQL

Snowflake + dbt — Continued

dbt is taught on top of Snowflake so candidates understand a realistic ELT workflow.

- dbt project structure, profiles/connections and development workflow
- Models: views, tables, incremental models and materializations
- ref() and source(); model dependencies and DAG thinking
- Staging, intermediate and marts layers; modular SQL transformation design
- Generic tests: not_null, unique, relationships and accepted_values; custom tests
- Seeds, snapshots and SCD Type 2 concepts
- Sources and freshness checks
- Macros, Jinja fundamentals, variables and reusable logic
- Documentation, lineage and model descriptions
- dbt run, test, build and selective execution
- Environment strategy and practical CI/CD concepts for dbt + Snowflake
- Debugging failed models, data-quality issues and production-style scenarios

MODULE 4 • PYTHON FOR DATA ENGINEERING

- Python syntax, variables, data types and operators
- Lists, tuples, sets and dictionaries with data-focused examples
- Conditions, loops, functions and reusable code
- String handling and common transformations
- Exception handling and defensive coding
- File handling: CSV/JSON basics and working with paths
- Modules, packages and virtual-environment concepts
- Pandas fundamentals for reading, filtering, transforming and validating datasets
- Writing small utilities used in ingestion, validation and automation workflows
- Python interview and practical coding exercises relevant to Data Engineering

MODULE 5 • DATABRICKS + PYSPARK

- Spark and Databricks architecture recap: driver, executors, cluster and distributed processing
- Databricks workspace, notebooks, compute and jobs
- PySpark DataFrames: create, read, select, filter, withColumn and transformations
- Spark SQL and switching between SQL and PySpark
- Joins, aggregations and window functions in PySpark
- Lazy evaluation, transformations/actions and narrow vs wide operations
- Reading/writing CSV, JSON, Parquet and Delta formats
- Delta Lake fundamentals, ACID transactions and MERGE
- Medallion architecture: Bronze, Silver and Gold
- Partitioning, caching, broadcast joins, skew awareness and performance basics
- Auto Loader / incremental ingestion concepts and production pipeline patterns
- Databricks-to-Snowflake integration patterns and practical scenarios

MODULE 6 • GITHUB & ENGINEERING WORKFLOW

- Git vs GitHub and why version control matters
- Repository, clone, status, add, commit, push and pull
- Branches, feature-branch workflow and merge concepts
- Pull requests, code review and resolving basic conflicts

- .gitignore, README and clean repository structure
- How SQL, Python, dbt and project code should be organized
- Basic CI/CD mindset for promoting changes across environments

2 End-to-End Real-World Projects

The projects bring the individual technologies together so candidates can explain an architecture end-to-end, not just isolated tools.

PROJECT 01	Azure → Snowflake → dbt Analytics Pipeline
Scenario	Build a batch-oriented cloud data pipeline from raw source files into curated Snowflake analytics models.
Hands-on flow	Source files → Azure storage/ingestion concepts → Snowflake stage & load → raw tables → dbt transformations → tests → curated marts → GitHub versioning.
Skills	Incremental loading, transformations, data quality, dimensional modelling concepts, orchestration, troubleshooting and deployment thinking.

PROJECT 02	Databricks → Delta → Snowflake Production-Style Pipeline
Scenario	Process larger datasets using Databricks/PySpark and deliver governed, analytics-ready data into Snowflake.
Hands-on flow	Raw data → PySpark transformations → Bronze/Silver/Gold Delta layers → incremental processing → Snowflake consumption → validation → GitHub workflow.
Skills	Schema handling, MERGE/incremental logic, Spark performance thinking, reconciliation, failure handling, environment strategy and end-to-end explanation.

What Candidates Practice Throughout

Hands-on Labs Queries, pipelines, transformations and debugging	Assignments Topic-wise exercises to reinforce each module
Assessments Regular checks on concepts and practical readiness	Interview Preparation Scenario questions and ready-to-explain project discussions
LMS Access Organized recordings and learning material	Project Readiness Architecture, implementation and troubleshooting mindset

Who This Program Is For

- Working professionals moving toward Data Engineering / Snowflake roles
- SQL or ETL developers who want a modern cloud data stack

- Data engineers who want stronger Snowflake, dbt and Databricks exposure
- Candidates preparing for practical Data Engineering interviews

D4DATA.in Snowflake-first • Practical • Project-driven

Curriculum sequence and depth may be adjusted slightly based on batch progress and practical requirements.