



AWS & databricks Training & Placement



aws Glue



kafka



Call: +91-8500002025

www.databrickstraining.in

MASTERING SNOWFLAKE & DBT

Data Engineering Excellence

Technologies: Snowflake • dbt • AWS • Azure • GCP • Apache Airflow • Kafka • PySpark • Python • SQL • Databricks

COURSE OVERVIEW

This comprehensive **120-hour course** covers everything you need to master Snowflake and dbt for enterprise data engineering. From fundamentals to advanced production patterns, you'll learn industry best practices used by leading data teams at Fortune 500 companies.

Who Should Attend: Data Engineers, Analytics Engineers, ETL Developers, Data Architects, and anyone wanting to build scalable, maintainable data pipelines.

Prerequisites: Basic SQL knowledge, familiarity with cloud platforms (AWS/Azure/GCP), and Python programming basics.

Training Format: Instructor-led with hands-on labs, real-world projects, and certification preparation.

MODULE 1: Data Warehousing & Snowflake Fundamentals

Duration: **3 Hours** | Level: **Beginner**

- What is Cloud Data Warehousing & Why Snowflake
- Snowflake vs Traditional Data Warehouses (Redshift, BigQuery, Synapse)
- Snowflake Architecture — Unique Design
 - Storage Layer — Centralized Cloud Storage


- Compute Layer — Virtual Warehouses
- Cloud Services Layer — Metadata & Optimization
- Separation of Storage & Compute — Key Advantage
- Snowflake Editions — Standard, Enterprise, Business Critical
- Snowflake on AWS, Azure & GCP — Multi-Cloud Support
- Snowflake Account Setup & Trial — Hands-On
- Snowflake Web UI — Snowsight Overview
- Snowflake Key Concepts
 - Databases, Schemas, Tables
 - Virtual Warehouses
 - Roles & Users
 - Stages


 **Labs:** Create Snowflake Trial Account, Explore Snowsight, Create Database & First SQL Queries

MODULE 2: Snowflake Storage & Data Loading

Duration: 4 Hours | **Level:** Beginner-Intermediate

- Snowflake Table Types: Permanent, Transient, Temporary, External Tables
- Snowflake File Formats: CSV, JSON, Parquet, ORC, Avro, XML
- Snowflake Stages
 - Internal Stages (User, Table, Named)
 - External Stages (S3, Azure Blob, GCS)
- Loading Data into Snowflake
 - COPY INTO Command
 - PUT Command
 - Snowpipe Auto-Ingest
 - INSERT Statement
- Loading Data from S3 to Snowflake — Step by Step
- Loading JSON & Semi-Structured Data
- COPY INTO Options: ON_ERROR, PURGE, FORCE
- Unloading Data — COPY INTO for Export
- Snowpipe Error Handling & Monitoring

 **Labs:** Load CSV from S3, Load JSON Data, Set Up Snowpipe Auto-Ingest


 **Quiz:** 15 Questions on Storage & Loading

MODULE 3: Snowflake SQL — Complete Fundamentals Guide

Duration: 5 Hours | **Level:** Intermediate

- Snowflake SQL — Standard SQL + Extensions
- DDL Commands: CREATE, ALTER, DROP, TRUNCATE
- DML Commands: INSERT, UPDATE, DELETE, MERGE
- Advanced Joins: Inner, Left, Right, Full, Cross, Lateral
- Aggregations: GROUP BY, HAVING, ROLLUP, CUBE
- Window Functions — Critical for Data Engineering
 - ROW_NUMBER, RANK, DENSE_RANK, NTILE
 - LEAD, LAG, FIRST_VALUE, LAST_VALUE

- Running Totals & Moving Averages
- Subqueries & CTEs (WITH Clause)
- Semi-Structured Data Functions
 - PARSE_JSON, FLATTEN, GET
 - OBJECT_CONSTRUCT, ARRAY operations
- String, Date & Timestamp Functions
- PIVOT & UNPIVOT Operations
- MERGE Statement — Upsert Pattern
- Stored Procedures & Snowflake Scripting
- UDFs & UDTFs — User Defined Functions & Table Functions

 **Labs:** Complex SQL with Window Functions, Semi-Structured Data Processing, MERGE Implementation

MODULE 4: Snowflake Virtual Warehouses & Performance

Duration: 4 Hours | **Level:** Intermediate

- Virtual Warehouse Architecture — Compute in Snowflake
- Warehouse Sizes: XS to 6XL and Custom Sizes
- Multi-Cluster Warehouses — Auto-Scale for Concurrency
- Warehouse Auto-Suspend & Auto-Resume Configuration
- Query Queuing — Concurrency Management
- Snowflake Credits — Understanding Cost Model
- Query Performance Optimization
 - Query Profile — Understand Execution Plans
 - Pruning & Partition Elimination
 - Spillage to Disk — Causes & Fixes
 - Cartesian Products & Exploding Joins
- Caching in Snowflake: Result, Local Disk, Remote Disk Cache
- Clustering Keys — When & How to Use
- Search Optimization Service — Point Lookups
- Materialized Views — Pre-Compute Expensive Queries
- Query Tags & Cost Attribution
- Resource Monitors — Credit Budget Control


 **Labs:** Query Profile Analysis, Resource Monitor Setup, Multi-Cluster Warehouse Configuration

MODULE 5: Snowflake Advanced Features

Duration: 5 Hours | **Level:** Intermediate-Advanced

- Time Travel — Query Historical Data (AT & BEFORE syntax)
- Restore Dropped Tables with UNDROP
- Time Travel Retention — 0 to 90 days based on edition
- Fail-Safe — 7-Day Disaster Recovery
- Zero-Copy Clone — Instant Database Cloning
- Data Sharing — Snowflake's Killer Feature
 - Share data without copying
 - Provider & Consumer accounts
 - Snowflake Marketplace — Pre-built datasets
- Snowflake Streams — Change Data Capture


- Standard Stream vs Append-Only Stream
- Stream offset & change tracking
- Snowflake Tasks — Scheduled SQL Execution
 - Standalone Tasks & Task Trees (DAG)
 - CRON & Interval scheduling
- Dynamic Tables — Declarative Data Pipelines
 - Target lag & freshness control
 - Incremental & full refresh modes
- Snowflake Alerts — Notify on Data Conditions

 **Labs:** Time Travel Recovery, Zero-Copy Clone, Streams + Tasks CDC Pipeline, Dynamic Tables

MODULE 6: Snowflake Data Governance & Security

Duration: 4 Hours | **Level:** Intermediate-Advanced

- Snowflake Access Control — RBAC Model
 - System Roles: ACCOUNTADMIN, SYSADMIN, SECURITYADMIN
 - Custom Roles & Role Hierarchy
 - Privileges: GRANT & REVOKE
- User Management — Create & Manage Users & Service Accounts
- Object-Level Security: Database, Schema, Table
- Column-Level Security
 - Dynamic Data Masking — Mask PII at Query Time
 - External Tokenization
- Row-Level Security — Row Access Policies
- Data Classification & Snowflake Tags
- Object Tagging & Business Metadata
- Access History & Audit — Who Accessed What
- Data Lineage & Impact Analysis
- Network Policy — IP Allowlist
- Encryption — AES-256, Always On
- Private Link — Secure Network Connectivity
- Compliance: SOC2, HIPAA, PCI-DSS, GDPR


 **Labs:** RBAC Setup, Dynamic Data Masking, Row Access Policies

MODULE 7: Snowflake Python & Developer Features

Duration: 4 Hours | **Level:** Intermediate

- Snowflake Python Connector — Connect from Python
- Execute Queries & Fetch Results
- Batch Insert & Bulk Load Operations
- Snowpark — Python, Java & Scala in Snowflake
 - Snowpark DataFrame API — Like PySpark
 - Reading & Writing Snowflake Tables
 - Transformations & Pushdown
- Snowpark UDFs — Python Functions in Snowflake
- Snowpark Stored Procedures
- Snowflake ML Functions


- Forecasting & Anomaly Detection
- Time series analysis
- Snowflake Notebooks — Interactive Python in Snowsight
- Snowflake Cortex — AI & LLM Functions
 - COMPLETE — Call LLM from SQL
 - SUMMARIZE, SENTIMENT, TRANSLATE, EXTRACT_ANSWER
- Connecting to BI Tools: Power BI, Tableau, Looker
- Snowflake REST API

 **Labs:** Snowpark ETL Pipeline, Snowflake Cortex AI Functions, BI Tool Connection

MODULE 8: dbt — Core Concepts & Setup

Duration: 4 Hours | **Level:** Intermediate

- What is dbt & The Analytics Engineering Workflow
- dbt vs Traditional ETL — ELT Approach
- dbt Core vs dbt Cloud — Differences & Use Cases
- dbt Architecture — How dbt Works
- dbt Project Structure: Models, Sources, Tests, Seeds, Macros
- Documentation & Auto-Generated Data Catalog
- Lineage Graph — Visual DAG of Data Flow
- dbt Installation & Setup
 - Install dbt Core: pip install dbt-snowflake
 - dbt Cloud Sign Up & Connect
 - Configure profiles.yml
- dbt Project Structure Deep Dive
- dbt Commands — Essential CLI
 - dbt run, dbt test, dbt docs
 - dbt debug, dbt seed, dbt snapshot


 **Labs:** Install dbt, Connect to Snowflake, Run First Model, Explore Project Structure

MODULE 9: dbt Models — Building Transformations

Duration: 5 Hours | **Level:** Intermediate

- What is a dbt Model — SQL SELECT Statement
- dbt Materializations: View, Table, Incremental, Ephemeral
- Incremental Models — Most Important Pattern
 - is_incremental() macro
 - unique_key & Merge strategy
 - Incremental strategies: append, merge, delete+insert
 - Late arriving data handling
- Model Layering — Best Practice Architecture
 - Staging Layer (stg_)
 - Intermediate Layer (int_)
 - Marts Layer (fct_ & dim_)
- Medallion Architecture with dbt
 - Bronze — Raw source data
 - Silver — Staging & intermediate


- Gold — Fact & dimension marts
- Referencing Models: ref() & source() Functions
- dbt Jinja — SQL + Python Templating
 - Variables, Macros, Conditional Logic, Loops
- dbt Tags — Label & Filter Models

 **Labs:** Build Staging/Intermediate/Mart Models, Incremental Model, Dynamic SQL

MODULE 10: dbt Testing & Data Quality

Duration: 4 Hours | **Level:** Intermediate

- Why Data Quality Testing Matters
- dbt Generic Tests — Built-In Tests
 - unique, not_null, accepted_values, relationships
- dbt Singular Tests — Custom SQL Assertions
- dbt Test Severity — warn vs error
- dbt Test Thresholds — Conditional Failures
- Great Expectations with dbt
- dbt-utils Package — Extra Tests & Macros
 - expression_is_true, recency, not_constant
- Elementary — dbt Native Data Observability
 - Anomaly detection on models
 - Data freshness monitoring
 - Column-level test coverage
- Re-Data — Data Monitoring for dbt
- Writing Tests for Different Model Types
 - Fact tables: grain, nulls, duplicates
 - Dimension tables: uniqueness, relationships
- Test Coverage Strategy


 **Labs:** Full Test Suite, Custom Singular Tests, Elementary Anomaly Detection

MODULE 11: dbt Advanced Features

Duration: 5 Hours | **Level:** Advanced

- dbt Macros — Reusable SQL Functions
 - Create custom macros with arguments & defaults
 - dispatch — Override macros per adapter
- dbt Packages — Community Extensions
 - dbt_utils, dbt_expectations, dbt_audit_helper, codegen
 - dbt_project_evaluator — Best practices checker
- dbt Hooks — Run SQL Before/After Models
 - pre-hook & post-hook
 - on-run-start & on-run-end
- dbt Snapshots — SCD Type 2 History
 - strategy: timestamp vs check
 - unique_key & valid_from/valid_to columns
- dbt Seeds — Load Static Data
 - CSV files as lookup tables


- dbt Analyses — Ad-Hoc SQL in Project
- dbt Exposures — Document Downstream Usage
- dbt Metrics — Define Business Metrics in YAML
- dbt Contracts — Enforce Column Types & Constraints
- dbt Versions — Model Versioning for Breaking Changes
- dbt Unit Tests — Test SQL Logic Without Running

 **Labs:** Macro Library, Snapshots SCD Type 2, Packages Integration

MODULE 12: dbt Documentation & Lineage

Duration: 3 Hours | **Level:** Intermediate

- dbt Documentation — Auto-Generated Data Catalog
- YAML Description Files
- Document Models, Columns, Sources, Tests
- dbt Docs Generate & Serve
- Lineage Graph — Visual DAG of Data Flow
 - View model dependencies
 - Filter by tag, folder, model
 - Identify upstream & downstream impact
- Column-Level Lineage
- Source Freshness — Monitor Raw Data Staleness
 - loaded_at_field configuration
 - Warn & error thresholds
- dbt Metadata in dbt Cloud
- Catalog.json — Machine-Readable Documentation
- Integrating with Data Catalog Tools: Atlan, Alation, DataHub

 **Labs:** Full Documentation, Lineage Graph Exploration, Impact Analysis

WHAT YOU'LL MASTER

- ✓ Build production-grade data pipelines on Snowflake
- ✓ Design scalable, cost-optimized data architectures
- ✓ Master advanced dbt for enterprise data transformation
- ✓ Implement real-time CDC pipelines with Streams & Tasks
- ✓ Optimize queries & manage performance at scale
- ✓ Implement comprehensive data governance & security
- ✓ Orchestrate complex workflows with Airflow
- ✓ Deploy CI/CD pipelines for data infrastructure
- ✓ Pass SnowPro Core & dbt Analytics Engineer certifications
- ✓ Design multi-tenant, enterprise data platforms

 **ENROLLMENT & CONTACT**

 **Website: databrickstraining.in**

 **Email: info@databrickstraining.in**

 **Phone: +91-8500002025**

Training & Placement Excellence | Since 2015

© 2024 Databricks Training. All rights reserved. | Last Updated: May 2026