

Snowflake Data Warehouse

Course Duration ~70 Hours · 9 Weeks · Live Online · Lab-Heavy · Portfolio of Production-Style Pipelines

Curriculum reviewed: July 2026 — Snowflake ships fast; this syllabus tracks the current platform

[Snowflake SQL](#)
[Snowpipe & Pipelines](#)
[Snowpark Python](#)
[Governance](#)
[Cost Engineering](#)
[Cortex AI](#)

In short: Go beyond writing SELECT statements. Most Snowflake training stops at the SQL warehouse — this program treats Snowflake as what enterprises actually standardise on: a data *platform*. Across 9 modules you ship ingestion pipelines with Snowpipe, incremental pipelines built twice (Streams + Tasks vs Dynamic Tables) and compared on complexity and cost, Snowpark Python jobs with a Streamlit app served from inside Snowflake, an Iceberg interoperability layer, a full governance build-out with masking and row access policies, a documented cost-rescue with credits saved, and Cortex AI analysis over governed data — closing with an end-to-end governed, cost-monitored platform capstone. One theme runs throughout: **cost is a design decision** — engineers who control credit consumption as well as write SQL are the ones who get noticed. Built for fresh IT graduates (working SQL is the only hard prerequisite) and for analysts, BI developers and legacy warehouse professionals, who have the shortest on-ramp. Aligned to the SnowPro Core certification path.

Duration	9 weeks · ~70 hours · live + hands-on	Format	Live online · lab-heavy · portfolio of production-style pipelines
Prerequisite	Working SQL. Python helpful for the Snowpark modules but taught from first principles. No prior cloud or warehousing experience required.	Curriculum	9 modules, from platform architecture to AI on governed data
Certification	Aligned to the SnowPro Core path; builds toward SnowPro Advanced: Data Engineer and the Snowpark specialty track	Ideal for	Fresh IT graduates and SQL-strong analysts, BI developers and legacy warehouse professionals moving to the cloud

Course Highlights

- Every module leads with what you build and ship, not what you "learn about"
- Platform, not just warehouse: pipelines, Snowpark Python, Iceberg, governance, AI
- Cost is a design decision: credit consumption engineered, measured and reported — a dedicated cost-rescue project
- Modern pipelines: Snowpipe, Streams & Tasks, and Dynamic Tables compared head-to-head
- Governance done properly: RBAC, masking, row access policies, lineage, secure sharing
- Cortex AI functions over governed structured and unstructured data
- dbt, Kafka, Airflow and Terraform where real teams actually use them

Hands-on Projects & Portfolio

- **End-to-End Ingestion Pipeline** — staged files → Snowpipe → Streams and Tasks → curated medallion layers
- **Dynamic Tables Rebuild** — the same pipeline rebuilt declaratively, compared on complexity and cost
- **Snowpark + Streamlit** — Python transformation job plus a Streamlit app served from inside Snowflake
- **Governance Build-Out** — role hierarchy, masking, row access policies and an audit trail for a simulated regulated dataset
- **Cost Optimisation Case Study** — profile a wasteful workload, fix it, and report credits saved with before/after evidence
- **Secure Data Share** — governed data shared to an external consumer with no data movement
- **Cortex AI Analysis** — AI over unstructured text sitting alongside governed structured data

Tools & Stack

Snowflake · SnowSQL · Snowsight · Snowpark (Python) · Snowpipe · Streams & Tasks · Dynamic Tables · Apache Iceberg · Cortex AI · Streamlit · dbt · Apache Kafka · Apache Airflow · AWS S3 · Azure Blob · GCS · Terraform · Git

Certification Alignment

- Syllabus mapped to the **SnowPro Core** certification path
- Builds toward **SnowPro Advanced: Data Engineer** and the Snowpark specialty track
- A lighter platform-level associate credential exists as an entry point
- Presented as preparation and alignment — no official partnership or accreditation implied, and no exam outcome guaranteed

Career Outcomes

Snowflake is one of the platforms enterprises standardise on rather than experiment with, which makes the skill unusually durable — and engineers who can control credit consumption as well as write SQL are the ones who get noticed. SQL-strong analysts and legacy warehouse developers have the shortest on-ramp. This course targets:

- Data Engineer · Snowflake Developer · Analytics Engineer · Data Warehouse Engineer
- Data Platform Engineer · BI Developer moving to the cloud · Data Architect · FinOps / platform cost analyst for data

Who This Course Is For

- Final-year students and fresh IT graduates targeting data engineering or analytics roles
- SQL-strong data analysts and BI developers moving into engineering
- Oracle, Teradata, SQL Server and legacy warehouse professionals migrating to the cloud
- ETL developers modernising onto ELT and the lakehouse
- Python and backend engineers joining data platform teams
- Data leads accountable for governance, compliance or a Snowflake bill

Module 1: Architecture That Explains Everything Else

7 Hours

YOU SHIP

- A working account layout — warehouses sized and suspended deliberately — plus an annotated architecture brief explaining why concurrency and cost behave differently here than in a traditional warehouse

TOPICS

- Separation of storage, compute and cloud services; virtual warehouses and multi-cluster scaling
- Micro-partitions and clustering; result and metadata caching
- Editions and cloud/region choices

Assessment: Quiz + Lab

Module 2: SQL, Modelling & Semi-Structured Data

8 Hours

YOU SHIP

- A dimensional model with medallion layering inside Snowflake — including JSON flattened straight from VARIANT with no pre-parsing, and a zero-copy clone used as a safe dev environment

TOPICS

- Snowflake SQL beyond the basics; window functions; the VARIANT type; querying and flattening JSON/Avro/Parquet
- Zero-copy cloning; Time Travel and Fail-safe
- Dimensional modelling and medallion layering inside Snowflake

Assessment: Quiz + Lab

Module 3: Data Loading & Ingestion

8 Hours

YOU SHIP

- A continuous ingestion pipeline — staged files through Snowpipe auto-ingest — with idempotent reloads, error handling, and file sizing chosen deliberately because it quietly determines your bill

TOPICS

- Stages (internal, external, named); COPY INTO; bulk vs continuous loading
- Snowpipe and auto-ingest; file sizing and format choices; streaming ingestion patterns and Kafka connectivity
- Error handling and idempotent reloads

Assessment: Quiz + Lab**Module 4: Pipelines & Transformation**

8 Hours

YOU SHIP

- The same incremental pipeline built twice — Streams + Tasks, then Dynamic Tables — compared honestly on complexity and cost, plus a dbt transformation layer

TOPICS

- Streams for change data capture; Tasks and task graphs
- Dynamic Tables for declarative incremental pipelines; stored procedures and UDFs; scheduling and dependency management
- dbt on Snowflake as the transformation layer

Assessment: Quiz + Lab**Module 5: Snowpark & Python in the Platform**

9 Hours

YOU SHIP

- A Snowpark Python transformation job that pushes logic to the data instead of exporting it, plus a Streamlit app served from inside Snowflake

TOPICS

- The Snowpark DataFrame API; Python UDFs and UDTFs; stored procedures in Python
- Snowpark-optimised warehouses; Streamlit apps inside Snowflake
- When Snowpark is the right choice versus pulling data out to Spark

Assessment: Quiz + Lab**Module 6: Open Table Formats & Interoperability**

7 Hours

YOU SHIP

- An Iceberg table layer readable from outside Snowflake, with a written trade-off decision: open lakehouse vs native tables for a given workload

TOPICS

- Apache Iceberg tables in Snowflake; managed vs externally managed; catalog interoperability
- External tables over object storage
- Architectural trade-offs of an open lakehouse versus native Snowflake tables

Assessment: Quiz + Lab**Module 7: Security & Governance**

8 Hours

YOU SHIP

- A full governance build-out for a simulated regulated dataset — role hierarchy, masking, row access policies, lineage and audit trail — plus a secure share to an external consumer with no data movement

TOPICS

- RBAC done properly; role hierarchies and least privilege; dynamic data masking; row access policies
- Tagging and classification; object lineage; access history and auditing; secure views; network policies
- Secure data sharing, the Marketplace and clean rooms

Assessment: Quiz + Lab

Module 8: Performance & Cost Engineering

8 Hours

YOU SHIP

- A cost optimisation case study: profile a wasteful workload, fix it, and report credits saved with before/after evidence — the artifact that gets Snowflake engineers noticed internally

TOPICS

- Reading the Query Profile; spillage; pruning and clustering decisions; search optimisation and materialised views
- Warehouse sizing; auto-suspend and auto-resume; workload isolation; resource monitors and budgets; credit attribution by team
- Newer workload-aware automatic compute options — when to hand sizing to the platform and when to control it yourself

Assessment: Quiz + Lab**Module 9: AI on Governed Data + Capstone**

7 Hours

YOU SHIP

- One end-to-end governed, cost-monitored Snowflake platform — designed, built and defended — including Cortex-powered analysis over unstructured text alongside governed structured data

TOPICS

- Cortex AI functions for summarisation, classification and sentiment directly in SQL
- Semantic views and natural-language querying over structured data; search over unstructured documents
- Governed data as the foundation for enterprise AI agents; capstone design, build and defense

Assessment: Capstone Project + Review