Snowflake Training – Course Content

## Day 1

**Module 1: Introduction to Snowflake**

* Overview of [Snowflake](https://chatgpt.com?q=Snowflake) as a cloud data platform
* Key features and benefits
* Editions and cloud providers (AWS, Azure, GCP)

**Module 2: Snowflake Architecture**

* Multi-cluster shared data architecture
* Separation of compute and storage
* Virtual warehouses
* Services layer overview

## Day 2

**Module 3: Snowflake Objects and SQL Commands**

* Understanding and working with:
  + Databases, Schemas, Tables (Permanent, Transient, Temporary)
  + [Views](https://chatgpt.com?q=Views): Standard and Materialized
  + [Stored Procedures](https://chatgpt.com?q=Stored%20Procedures) using JavaScript and SQL
* DDL, DML, and Query constructs
* User roles and access control

**Module 4: Data Loading in Snowflake Using AWS S3**

* External stages and file formats
* Loading data using COPY INTO
* Automating ingestion with [Snowpipe](https://chatgpt.com?q=Snowpipe)
* Best practices for bulk vs. continuous loading

**Module 5: Understanding Stages in Snowflake**

* Types of stages: User, Table, Internal, External (S3)
* File format options and compression
* Secure file transfer with PUT and GET commands

## Day 3

**Module 6: Snowflake Caching Mechanism**

* Result cache
* Metadata cache
* Data cache
* How caching improves performance and cost-efficiency

**Module 7: Snowflake Storage & Optimization**

* Columnar storage and [Micro-partitioning](https://chatgpt.com?q=Micro-partitioning)
* Automatic clustering vs. Manual clustering
* Storage cost considerations and optimizations

**Module 8: Performance Tuning in Snowflake**

* Query profiling and execution plan analysis
* Optimizing warehouse size and scaling
* Query history and monitoring
* Techniques for reducing compute and storage cost

## Day 4

**Module 9: Working with Various Data Formats**

* Ingesting and querying [JSON](https://chatgpt.com?q=JSON), [XML](https://chatgpt.com?q=XML), [Parquet](https://chatgpt.com?q=Parquet), Avro
* Semi-structured data support using VARIANT
* Flattening and transforming nested data

**Module 10: Tasks and Streams**

* Use cases for [Tasks](https://chatgpt.com?q=Tasks) and [Streams](https://chatgpt.com?q=Streams)
* Creating continuous data pipelines
* Change data capture (CDC)
* Orchestration of automated workflows

## Day 5

**Module 11: Advanced Data Replication and Failover**

* Cross-Region and Cross-Cloud Replication
* Ensuring High Availability and Disaster Recovery
* Automating Failover and Recovery Processes

**Module 12: Designing and Implementing a Complete Data Pipeline**

* Real-World Use Case: End-to-End Data Warehousing Solution
* Presentation and Peer Review