

7-Day Snowflake + Azure Databricks + Snowpark API Training Plan

Day 1: Introduction to Snowflake, Azure Databricks, and Snowpark API

Objective: Understand Snowflake architecture, set up Azure-Snowflake connectivity, and introduce Snowpark API for Databricks.

Topics

- Snowflake Architecture: Storage, Compute, and Cloud Services
- Azure Databricks Overview and Integration with Snowflake
- **Snowpark API for Python/Scala:** Overview and Setup
- Azure Data Lake Storage (ADLS) and Snowflake Stages
- Snowflake Web UI, SnowSQL CLI, and Azure Active Directory (AAD) Integration
- Role-Based Access Control (RBAC) in Snowflake and Azure

Labs

1. **Set Up Snowflake Trial Account** and connect to Azure AD for SSO.
2. **Create Databases, Schemas, and Tables** in Snowflake using DDL.
3. **Connect Azure Databricks to Snowflake** using the Snowflake Connector for Spark.
4. **Mount Azure Data Lake Storage (ADLS) in Databricks** and load data into Snowflake.
5. **Set Up Roles and Privileges** in Snowflake for Databricks users.
6. **Explore Snowflake Marketplace** and load a sample dataset.
7. **Write a Snowpark Python Script** in Databricks to query Snowflake data.

Case Study

- **Scenario:** Migrate an on-premises SQL Server database to Snowflake via Azure Databricks.
 - **Task:** Extract data from SQL Server, transform it in Databricks using Snowpark, and load it into Snowflake.
-

Day 2: Data Ingestion and Transformation with Snowpark API

Objective: Ingest and transform data from Azure services into Snowflake using Databricks, PySpark, and Snowpark API.

Topics

- Data Ingestion: COPY INTO, Snowpipe, and Azure Synapse Integration
- Azure Data Factory (ADF) and Snowflake Integration
- PySpark Transformations in Databricks for Snowflake
- **Snowpark API for Data Transformation**
- Semi-Structured Data (JSON, Parquet, Delta Lake) in Snowflake
- File Formats and Stages in Snowflake

Labs

1. **Load Data from ADLS to Snowflake** using COPY INTO and external stages.
2. **Set Up Snowpipe with Azure Event Grid** for real-time data ingestion.
3. **Transform Data in Databricks** using PySpark and Snowpark API.
4. **Ingest Delta Lake Tables** from Databricks into Snowflake using Snowpark.
5. **Use Azure Data Factory** to orchestrate data movement between Databricks and Snowflake.
6. **Query Semi-Structured Data** (JSON, Parquet) in Snowflake using Snowpark.
7. **Write a Snowpark UDF** to cleanse and transform data.

Case Study

- **Scenario:** Build a real-time log analytics pipeline using Azure Databricks, Snowpark, and Snowflake.
 - **Task:** Ingest JSON logs from ADLS, transform them in Databricks using Snowpark, and load them into Snowflake for analysis.
-

Day 3: Performance Optimization and Snowpipe with Snowpark

Objective: Optimize Snowflake performance, scale warehouses, and implement Snowpipe with Azure services and Snowpark.

Topics

- Virtual Warehouse Scaling and Clustering in Snowflake
- Caching and Query Optimization
- Snowpipe with Azure Blob Storage and Event Grid
- **Snowpark for Performance Tuning**
- Resource Monitors and Cost Management
- Azure Functions for Snowflake Automation

Labs

1. **Scale Virtual Warehouses** and test performance with large datasets using Snowpark.
2. **Set Up Snowpipe with Azure Blob Storage** and monitor using COPY_HISTORY.
3. **Automate Snowpipe with Azure Functions** to trigger data loads.

4. **Optimize Queries** using clustering keys and materialized views with Snowpark.
5. **Set Up Resource Monitors** to track credit usage and set alerts.
6. **Use Snowpark to Analyze Query Performance** and suggest optimizations.
7. **Create a Multi-Cluster Warehouse** and test concurrency with Snowpark.

Case Study

- **Scenario:** Optimize a slow-running analytics dashboard by tuning Snowflake warehouses and queries using Snowpark.
 - **Task:** Identify bottlenecks, apply clustering, and automate scaling with Azure Functions and Snowpark.
-

Day 4: Security, Governance, and CDC with Snowpark

Objective: Implement security, governance, and Change Data Capture (CDC) using Snowflake, Databricks, Azure services, and Snowpark.

Topics

- Role-Based Access Control (RBAC) and Azure AD Integration
- Column-Level and Row-Level Security in Snowflake
- Change Data Capture (CDC) with Streams and Tasks
- **Snowpark for Secure Data Processing**
- Azure Key Vault Integration for Secrets Management
- Data Masking and Encryption in Snowflake

Labs

1. **Set Up Azure AD Integration** for Snowflake authentication.
2. **Implement Column-Level Security** (data masking) in Snowflake using Snowpark.
3. **Create Secure Views** for row-level security with Snowpark.
4. **Set Up CDC with Streams and Tasks** to track changes in Databricks tables.
5. **Use Azure Key Vault** to manage Snowflake credentials securely in Snowpark.
6. **Automate Role Assignments** using Azure Logic Apps and Snowpark.
7. **Write a Snowpark Stored Procedure** to apply dynamic data masking.

Case Study

- **Scenario:** Secure a multi-tenant healthcare analytics platform in Snowflake using Snowpark.
 - **Task:** Implement RBAC, data masking, and CDC to track patient record updates with Snowpark.
-

Day 5: ELT Pipelines with Databricks, Snowflake, and Snowpark API

Objective: Build ELT pipelines using Azure Databricks, Snowflake, Azure Data Factory, and Snowpark API.

Topics

- ELT Pipeline Design with Databricks, Snowflake, and Snowpark
- Dynamic Tables and Hybrid Tables in Snowflake
- Stored Procedures and UDFs in Snowflake using Snowpark
- Azure Data Factory for Orchestration
- Zero-Copy Cloning and Time Travel
- **Snowpark for ELT Automation**

Labs

1. **Build an ELT Pipeline** using Databricks, Snowflake, Azure Data Factory, and Snowpark.
2. **Create Dynamic Tables** in Snowflake for incremental data processing with Snowpark.
3. **Write Stored Procedures** in Snowflake using Snowpark to automate transformations.
4. **Use Zero-Copy Cloning** to create test environments with Snowpark.
5. **Set Up Time Travel** to query historical data using Snowpark.
6. **Orchestrate the Pipeline** using Azure Data Factory and Snowpark.
7. **Develop a Snowpark Application** to monitor and log ELT pipeline execution.

Case Study

- **Scenario:** Design an ELT pipeline for a retail data warehouse using Snowpark.
 - **Task:** Extract sales data from Azure SQL, transform it in Databricks using Snowpark, and load it into Snowflake.
-

Day 6: Advanced Analytics and Integration with Power BI and Snowpark

Objective: Integrate Snowflake with Azure Databricks, Power BI, and Snowpark API for advanced analytics and machine learning.

Topics

- **Snowpark for Python and Scala** in Databricks
- Power BI Integration with Snowflake
- Machine Learning with Snowpark and Databricks
- Azure Synapse Link for Snowflake
- Real-Time Analytics with Azure Databricks, Snowflake, and Snowpark
- **Snowpark UDFs for ML and Analytics**

Labs

1. **Use Snowpark in Databricks** to query and transform Snowflake data.
2. **Build a Power BI Dashboard** connected to Snowflake using DirectQuery.
3. **Train a Machine Learning Model** in Databricks using Snowpark and Snowflake data.
4. **Set Up Azure Synapse Link** to query Snowflake data in Synapse.
5. **Automate ML Model Deployment** using Azure ML and Snowpark.
6. **Create a Real-Time Analytics Dashboard** in Power BI using Snowpark.
7. **Develop a Snowpark UDF** for predictive analytics (e.g., sales forecasting).

Case Study

- **Scenario:** Build a real-time sales analytics dashboard using Power BI, Snowflake, and Snowpark.
 - **Task:** Use Snowpark to preprocess data, train a forecasting model in Databricks, and visualize results in Power BI.
-

Day 7: Data Sharing, Disaster Recovery, and Capstone Project with Snowpark

Objective: Implement data sharing, disaster recovery, and a **capstone project** using Snowpark API to consolidate all skills.

Topics

- Secure Data Sharing with Azure Databricks and Snowpark
- Database Replication and Failover in Snowflake
- Snowflake Marketplace for Data Exchange
- Performance Tuning with Clustering and Search Optimization
- **Snowpark for Data Sharing and Replication**
- Capstone Project: End-to-End ELT Pipeline with Snowpark

Labs

1. **Set Up Secure Data Sharing** between Snowflake accounts using Snowpark.
2. **Configure Database Replication** for disaster recovery with Snowpark.
3. **Publish a Dataset to Snowflake Marketplace** (simulated) using Snowpark.
4. **Tune Performance** using clustering and search optimization with Snowpark.
5. **Capstone Project:**
 - Design an **end-to-end ELT pipeline** using Databricks, Snowflake, Azure services, and Snowpark.
 - Ingest data from Azure SQL/ADLS, transform it in Databricks using Snowpark, and load it into Snowflake.
 - Implement **security, performance tuning, monitoring, and automation** with Snowpark.

- Share insights via Power BI and replicate data for disaster recovery.

Case Study

- **Scenario:** Build a scalable, secure ELT pipeline for a healthcare analytics platform using Snowpark.
- **Task:**
 - Ingest patient data from Azure SQL and ADLS.
 - Transform and enrich data using Databricks and Snowpark.
 - Secure the environment with RBAC, masking, and row-level security.
 - Automate the pipeline with Azure Data Factory and monitor performance using Snowpark.
 - Share insights via Power BI and replicate data for disaster recovery.

Day	Focus Area	Key Labs	Case Study
1	Snowflake + Azure + Snowpark Setup	Snowflake-Azure AD integration, Databricks-Snowflake connection, ADLS data loading, Snowpark queries	SQL Server migration to Snowflake
2	Data Ingestion & Snowpark Transformation	COPY INTO, Snowpipe with Azure Event Grid, PySpark + Snowpark transformations, Delta Lake integration	Real-time log analytics pipeline
3	Performance & Snowpipe with Snowpark	Warehouse scaling, Snowpipe with Azure Blob, Azure Functions automation, Snowpark query optimization	Optimize analytics dashboard
4	Security & CDC with Snowpark	Azure AD SSO, column/row-level security with Snowpark, CDC with Streams, Azure Key Vault integration	Secure healthcare analytics platform
5	ELT Pipelines with Snowpark	ELT pipeline with ADF, Dynamic Tables, stored procedures, Zero-Copy Cloning, Snowpark automation	Retail data warehouse pipeline
6	Advanced Analytics & Snowpark	Snowpark in Databricks, Power BI dashboard, ML with Snowpark, Azure Synapse Link	Real-time sales analytics dashboard
7	Data Sharing & Capstone with Snowpark	Secure data sharing, database replication, Marketplace, capstone ELT pipeline with Snowpark	Healthcare analytics platform