DUAL-PLATFORM ENTERPRISE DATA & ANALYTICS PROGRAM

Microsoft Fabric + Databricks Combo Program

Comprehensive, career-accelerating mastery of both Microsoft Fabric (OneLake, Data Factory, Synapse Data Engineering & Warehouse, Real-Time Intelligence, Power BI Direct Lake) and Azure Databricks (Lakehouse, Delta Lake, PySpark, Unity Catalog, and Solution Architecture).

COURSE OVERVIEW

Modern enterprises increasingly operate hybrid Lakehouse architectures combining **Databricks** for heavy distributed Spark engineering, Delta Lake optimization, and ML/AI pipelines with **Microsoft Fabric** for unified SaaS analytics, OneLake storage, Data Factory orchestration, and Direct Lake Power BI reporting. This dual-track program—mapped to **Microsoft Certified: Fabric Data Engineer Associate (DP-700)** and **Databricks Certified Data Engineer / Solution Architect** paths—equips professionals to design, integrate, govern, and optimize both platforms end to end.

KEY LEARNING OUTCOMES

- **Master End-to-End Microsoft Fabric:** Build unified OneLake architectures using Fabric Lakehouses, Data Warehouses, Data Factory pipelines, Dataflows Gen2, and Real-Time Intelligence (Eventstreams & KQL).
- **Master Enterprise Databricks & Spark Engineering:** Architect Bronze-Silver-Gold Medallion platforms using Delta Lake, Delta Live Tables (DLT), Apache Spark tuning, and Unity Catalog.
- **Architect Fabric + Databricks Interoperability:** Connect OneLake Shortcuts with Unity Catalog external locations and Databricks Delta tables for zero-copy cross-platform analytics.
- **Deliver High-Speed Direct Lake BI & Governance:** Serve Power BI semantic models directly over Delta/Parquet files in Direct Lake mode with enterprise RBAC, Purview lineage, and FinOps controls.

PART I — MICROSOFT FABRIC DATA ENGINEERING TRACK

Module 1: Microsoft Fabric Architecture, Capacities & OneLake Foundations

- **SaaS Unified Analytics Architecture:** Fabric workspaces, capacity licensing (F-SKUs), compute engines, and tenant governance settings.
- **OneLake & Delta-Parquet Standardization:** Single-tenant data lake architecture, OneLake File Explorer, and open Delta Lake format compatibility across engines.
- **OneLake Shortcuts & Mirroring:** Zero-copy federation with Azure ADLS Gen2, AWS S3, Google Cloud Storage, Dataverse, and Database Mirroring (Azure SQL, Snowflake, Cosmos DB).

> **Hands-On Lab:** Configure a Fabric workspace, create a Lakehouse, set up OneLake Shortcuts to external cloud storage, and explore mirrored operational data.

Module 2: Data Ingestion & Orchestration — Data Factory & Dataflows Gen2

- **Fabric Data Pipelines:** Copy Activity throughput optimization, Foreach/If-Condition control flows, parameterization, and metadata-driven ingestion frameworks.
- **Dataflows Gen2 (Power Query at Scale):** Low-code visual transformations, query folding, incremental refresh, and Lakehouse/Warehouse output destinations.
- **Scheduling, Alerts & Data Activator:** Event-driven pipeline triggers and automated threshold alerts with Reflex/Data Activator.

> **Hands-On Lab:** Build a metadata-driven Data Factory pipeline ingesting multi-table ERP sources into Bronze OneLake tables and transform records via Dataflows Gen2.

Module 3: Synapse Data Engineering — Fabric Lakehouse & Spark Notebooks

- **Fabric Spark Compute & Starter Pools:** Live pools, custom Spark environments, library management, and high-concurrency notebook sessions.
- **PySpark & Spark SQL Transformations:** Implementing Bronze-Silver-Gold Medallion layers, V-Order write optimization, and automated table maintenance (OPTIMIZE/VACUUM).
- **Structured Streaming in Fabric:** Continuous micro-batch ingestion and stateful stream processing into Delta tables.

> **Hands-On Lab:** Develop parameterized PySpark notebooks in Fabric to cleanse raw JSON/CSV feeds into V-Order optimized Silver and Gold Delta tables.

Module 4: Synapse Data Warehouse & Direct Lake Power BI Serving

- **Fabric Data Warehouse Internals:** Distributed T-SQL query engine, multi-table ACID transactions, zero-copy clones, and cross-database querying.
- **Dimensional Modeling & SQL Analytics Endpoint:** Designing star schemas, views, stored procedures, and T-SQL COPY INTO ingestion.
- **Power BI Direct Lake Mode:** Eliminating Import/DirectQuery trade-offs by querying OneLake Delta files directly with automatic framing and fallback rules.

> **Hands-On Lab:** Model a star-schema Gold Data Warehouse in Fabric, configure row/column security, and publish a sub-second Direct Lake Power BI semantic model.

Module 5: Real-Time Intelligence, Security & Fabric CI/CD

- **Real-Time Intelligence (Eventstreams & KQL):** Ingesting streaming telemetry via Eventstreams, querying high-velocity logs in Eventhouse (KQL Database), and real-time dashboards.
- **Fabric Security & Microsoft Purview:** Workspace roles, OneLake Security (folder/table RBAC), sensitivity labels, and end-to-end lineage in Microsoft Purview.
- **Git Integration & Deployment Pipelines:** Connecting workspaces to Azure DevOps/GitHub and promoting artifacts across Dev, Test, and Prod stages.

> **Hands-On Lab:** Stream simulated IoT telemetry into an Eventhouse KQL database, apply workspace RBAC, and promote a Fabric solution across Dev and Prod via Git.

PART II — DATABRICKS DATA ENGINEERING & SOLUTION ARCHITECTURE TRACK

Module 6: Databricks Lakehouse Architecture, Clusters & Multi-Workspace Design

- **Databricks Control & Compute Plane:** Workspace architecture, All-Purpose vs. Job clusters, Photon acceleration, Serverless compute, and Databricks SQL Warehouses.
- **Enterprise Multi-Workspace Strategy:** Environment isolation across Dev, QA, and Production workspaces with secure VNet injection and secret scopes.
- **Medallion Architecture at Scale:** Designing enterprise Bronze, Silver, and Gold data lifecycles across cloud object storage.

> **Hands-On Lab:** Provision Databricks compute policies, configure Azure Key Vault / Databricks Secret Scopes, and set up isolated Dev/Prod workspace structures.

Module 7: Delta Lake Internals & Lakehouse Storage Optimization

- **Delta Transaction Log (_delta_log):** ACID guarantees, optimistic concurrency control, schema enforcement, schema evolution, and Time Travel queries.
- **Layout & File Compaction Tuning:** OPTIMIZE bin-packing, Z-Ordering, Liquid Clustering, Auto Optimize, and VACUUM retention management.
- **Change Data Feed (CDF):** Capturing row-level inserts, updates, and deletes for efficient downstream incremental propagation.

> **Hands-On Lab:** Benchmark query acceleration on multi-million-row Delta tables using Liquid Clustering, Z-Ordering, and Change Data Feed incremental merges.

Module 8: Production Pipelines — Auto Loader, Delta Live Tables & Workflows

- **Incremental Ingestion with Auto Loader:** CloudFiles schema inference, schema evolution rescue columns, and scalable file notification ingestion.
- **Delta Live Tables (DLT):** Declarative ETL pipelines, Streaming Tables, Materialized Views, Data Quality Expectations (WARN, DROP, FAIL), and SCD Type 1/2 via APPLY CHANGES.
- **Databricks Workflows & Monitoring:** Multi-task DAG orchestration, job repair/retries, cluster event logs, and failure alerting.

> **Hands-On Lab:** Build an end-to-end DLT pipeline ingesting streaming files via Auto Loader, enforcing data quality expectations, and maintaining SCD Type 2 history.

Module 9: Unity Catalog Governance, Security & FinOps Cost Optimization

- **Unity Catalog 3-Level Namespace:** Metastores, Catalogs, Schemas, Managed vs. External tables, Volumes, and automated column-level data lineage.
- **Fine-Grained Access Control:** Privilege inheritance, Dynamic Views, Row-Level Filters, Column Masks, and audit log analysis.
- **Cluster & Cost FinOps:** Right-sizing instance types, autoscaling, Spot fallback policies, serverless SQL budgets, and system table billing queries.

> **Hands-On Lab:** Configure a Unity Catalog hierarchy, implement dynamic PII column masking, trace column lineage, and build a DBU cost-monitoring dashboard.

Module 10: Apache Spark Performance Tuning & Databricks Asset Bundles (CI/CD)

- **Spark Execution & Query Tuning:** Spark UI DAG analysis, Adaptive Query Execution (AQE), Broadcast Hash Joins, caching strategies, and shuffle reduction.
- **Solving Data Skew & Spill:** Key salting techniques, AQE skew join optimization, repartition vs. coalesce, and memory/disk spill elimination.
- **CI/CD with Databricks Asset Bundles (DABs):** Packaging notebooks, DLT pipelines, and Workflows as YAML code for automated GitHub Actions / Azure DevOps deployment.

> **Hands-On Lab:** Diagnose and resolve a heavily skewed multi-terabyte join in Spark UI using salting and AQE, then deploy the pipeline via Databricks Asset Bundles.

PART III — HYBRID FABRIC + DATABRICKS INTEGRATION & CAPSTONE PROJECT

Module 11: Unified Fabric + Databricks Enterprise Architecture & Interoperability

- **Zero-Copy Cross-Platform Integration:** Exposing Databricks Unity Catalog Delta tables directly inside Microsoft Fabric OneLake via Shortcuts and Catalog Mirroring.
- **Workload Placement Strategy:** Deciding when to use Databricks Spark/ML vs. Fabric Warehouse/Data Factory/Real-Time Intelligence in enterprise reference architectures.
- **End-to-End Enterprise Case Studies:** Reference blueprints for Banking, Retail, Healthcare, and IoT streaming platforms.

> **Hands-On Lab:** Connect a Databricks Gold Delta table into a Microsoft Fabric Lakehouse via OneLake Shortcut and serve a live Power BI Direct Lake report without data duplication.

Component	Enterprise Deliverable & Scope
Dual-Platform Capstone Project: Hybrid Enterprise Lakehouse	Architect and implement an end-to-end production Lakehouse spanning Databricks and Microsoft Fabric. Ingest batch and streaming data into Databricks using Auto Loader and Delta Live Tables, enforce Unity Catalog row/column security, expose curated Gold Delta tables into Microsoft Fabric OneLake via Shortcuts, build a Fabric Warehouse dimensional model, and deliver a Power BI Direct Lake executive dashboard.

Component	Enterprise Deliverable & Scope
Certification & Solution Architect Interview Prep	Structured preparation for **Microsoft DP-700 (Fabric Data Engineer Associate)** and **Databricks Certified Data Engineer / Solution Architect** interviews, including architecture trade-off drills, resume optimization, and 1-on-1 mock interviews.

> **Enrollment & Support:** All learners receive lifetime access to live session recordings, Fabric & Databricks code repositories, lab guides, and mentor support. Visit www.nityacloudtech.com for enrollment details.