Data Engineering on Microsoft Azure

Content

Module 1: Explore Compute and Storage Options for Data Engineering Workloads

- **Lessons**:
    - Introduction to Azure Synapse Analytics
    - Azure Databricks and Delta Lake Architecture
    - Azure Data Lake Storage
    - Azure Stream Analytics for Data Streams
- **Lab**:
    - Combine streaming and batch processing
    - Organize the data lake
    - Index data lake storage for queries
- **After Completion**:
    - Understanding of Azure Synapse, Databricks, and Azure Stream Analytics.

Module 2: Design and Implement the Serving Layer

- **Lessons**:
    - Design multidimensional schema
    - Code-free transformations with Azure Data Factory
- **Lab**:
    - Design a star schema
    - Populate slowly changing dimensions
- **After Completion**:
    - Design star schemas and manage incremental data loads.

Module 3: Data Engineering Considerations for Source Files

- **Lessons**:
    - Modern Data Warehouse design with Azure Synapse
    - Secure data storage in Azure
- **Lab**:
    - Manage and secure files in Azure Data Lake
- **After Completion**:
    - Secure and design modern data warehouses using Azure Synapse.

Module 4: Run Interactive Queries Using Serverless SQL Pools

- **Lessons**:
    - Explore serverless SQL pools
    - Query data from the lake using T-SQL
- **Lab**:
    - Query Parquet and CSV files
    - Manage data security with RBAC and ACL
- **After Completion**:
    - Perform queries and manage data security with serverless SQL pools.

Module 5: Explore, Transform, and Load Data into the Data Warehouse Using Apache Spark

- **Lessons**:
    - Data transformation with Apache Spark in Synapse Analytics
    - Integrate SQL and Spark pools
- **Lab**:
    - Ingest and transform data using Apache Spark notebooks
- **After Completion**:
    - Use Apache Spark for data engineering tasks.

Module 6: Data Exploration and Transformation in Azure Databricks

- **Lessons**:
    - Explore and transform data using DataFrames in Azure Databricks
- **Lab**:
    - Filter, aggregate, and manipulate data using DataFrames
- **After Completion**:
    - Proficiency in working with DataFrames in Azure Databricks.

Module 7: Ingest and Load Data into the Data Warehouse

- **Lessons**:
    - Use PolyBase and COPY for data ingestion
- **Lab**:
    - Perform large-scale data ingestion with Azure Synapse Pipelines
- **After Completion**:
    - Master best practices for data loading in Azure Synapse.

Module 8: Transform Data with Azure Data Factory or Azure Synapse Pipelines

- **Lessons**:
    - Build integration pipelines
    - Code-free transformations at scale
- **Lab**:
    - Create data pipelines and mapping data flows
- **After Completion**:
    - Implement data integration and transformations.

Module 9: Orchestrate Data Movement and Transformation in Azure Synapse Pipelines

- **Lessons**:
    - Orchestrate data workflows with Synapse Pipelines
- **Lab**:
    - Integrate data from notebooks with Synapse Pipelines
- **After Completion**:
    - Automate and orchestrate data workflows using Synapse Pipelines.

Module 10: Optimize Query Performance with Dedicated SQL Pools

- **Lessons**:
    - Optimize query performance using Azure Synapse SQL pools
- **Lab**:
    - Improve query performance through optimization techniques
- **After Completion**:
    - Proficiency in optimizing Azure Synapse SQL pool queries.

Module 11: Analyze and Optimize Data Warehouse Storage

- **Lessons**:
    - Analyze table space usage and optimize storage
- **Lab**:
    - Explore storage optimization and use materialized views
- **After Completion**:
    - Optimize Azure Synapse storage for better performance.

Module 12: Support Hybrid Transactional Analytical Processing (HTAP) with Azure Synapse Link

- **Lessons**:
  - o Enable Synapse Link with Azure Cosmos DB
- **Lab**:
  - o Query Cosmos DB data using Apache Spark and SQL pools
- **After Completion**:
  - o Implement HTAP using Azure Synapse Link and Cosmos DB.

Module 13: End-to-End Security with Azure Synapse Analytics

- **Lessons**:
  - o Implement security using Active Directory and Key Vault
- **Lab**:
  - o Secure Synapse infrastructure and manage secrets
- **After Completion**:
  - o Implement comprehensive security solutions in Azure Synapse.

Module 14: Real-Time Stream Processing with Stream Analytics

- **Lessons**:
  - o Process streaming data using Azure Stream Analytics
- **Lab**:
  - o Ingest and process data streams from Event Hubs
- **After Completion**:
  - o Build real-time data stream processing solutions.

Module 15: Create a Stream Processing Solution with Event Hubs and Azure Databricks

- **Lessons**:
  - o Process streaming data with Databricks Structured Streaming
- **Lab**:
  - o Implement streaming solutions with Event Hubs and Databricks
- **After Completion**:
  - o Build scalable streaming solutions using Azure Databricks and Event Hubs.