

Mastering Data Analytics and Data Governance with Azure Synapse Analytics

Course Description

This 4-day immersive training program provides participants with a comprehensive understanding of Azure Synapse Analytics for data analytics and governance. The course covers foundational concepts, hands-on implementation, advanced analytics, and governance best practices. Learners will gain practical skills to design, secure, and optimize enterprise-grade analytics solutions.

Duration

4 Days (32 hours total)

Pre-requisites

- Basic knowledge of databases and SQL
- Familiarity with cloud computing concepts
- Exposure to Microsoft Azure services (preferred)
- Understanding of data warehousing principles

Learning Objectives

By the end of this course, participants will:

- Understand Azure Synapse architecture and capabilities
- Implement data ingestion and transformation pipelines
- Build analytical models and integrate with visualization tools
- Apply governance, compliance, and security practices
- Optimize Synapse workloads for performance and cost efficiency

Content Coverage

Day 1 – Foundations of Data Analytics & Synapse Overview

Module 1: Introduction to Data Analytics

- Importance of data-driven decision making

- Types of analytics: descriptive, diagnostic, predictive, prescriptive
- Role of data governance in analytics lifecycle
- Key challenges in enterprise data management
- Benefits of cloud-based analytics solutions

Module 2: Azure Synapse Analytics Overview

- Synapse architecture and core components
- Dedicated SQL pools vs serverless SQL pools
- Integration with Azure Data Lake Storage
- Synapse pipelines for orchestration
- Use cases and industry applications

Module 3: Setting Up Synapse Workspace

- Creating and configuring a Synapse workspace
- Managing linked services and datasets
- Workspace security and access management basics
- Configuring networking and firewall rules
- Monitoring workspace health and usage

Day 2 – Data Ingestion, Modeling & Analytics

Module 4: Data Ingestion & Preparation

- Connecting to structured and unstructured data sources
- Using Synapse pipelines for ETL/ELT workflows
- Data transformation with Data Flows
- Handling batch vs streaming data ingestion
- Best practices for data staging and cleansing

Module 5: Querying & Modeling Data

- Writing queries in dedicated and serverless SQL pools
- Designing star and snowflake schemas
- Partitioning and indexing strategies
- Creating materialized views for performance
- Managing metadata and schema evolution

Module 6: Building Analytical Models

- Dimensional modeling concepts
- Creating fact and dimension tables
- Implementing slowly changing dimensions (SCD)
- Using Synapse Studio for model development
- Validating and testing analytical models

Day 3 – Advanced Analytics & Visualization

Module 7: Advanced Querying & Performance

- Query optimization techniques
- Using statistics and query hints
- Handling large datasets efficiently
- Parallel processing in Synapse
- Monitoring query performance

Module 8: Visualization & Insights

- Integrating Synapse with Power BI
- Creating interactive dashboards and reports
- Real-time analytics scenarios
- Embedding analytics into applications
- Best practices for visualization design

Module 9: Machine Learning Integration

- Connecting Synapse with Azure Machine Learning
- Preparing data for ML models
- Running predictive analytics in Synapse
- Operationalizing ML pipelines
- Case studies of ML-driven analytics

Day 4 – Governance, Security & Optimization

Module 10: Data Governance Fundamentals

- Principles of data governance
- Metadata management and data cataloging
- Role-based access control (RBAC)
- Data lineage and auditing
- Policies for data quality and stewardship

Module 11: Security & Compliance in Synapse

- Data encryption at rest and in transit
- Dynamic data masking and row-level security
- Auditing and monitoring access logs
- Compliance with GDPR, HIPAA, and other standards
- Implementing secure data sharing

Module 12: Performance Optimization & Best Practices

- Indexing and partitioning strategies revisited
- Cost optimization in Synapse workloads
- Scaling dedicated SQL pools
- Monitoring and troubleshooting performance issues
- Best practices for enterprise adoption

Module 13: Capstone Project & Wrap-Up

- Hands-on project: building a governed analytics pipeline in Synapse
- Applying governance and security controls
- Performance tuning and optimization exercises
- Review of key concepts and lessons learned
- Q&A, feedback, and advanced learning roadmap