Building End to End solution using Azure Synapse Analytics

Course Description: This course provides an in-depth understanding of developing end-to-end solutions using Azure Synapse Analytics. Participants will learn about data integration, scalability, performance optimization, pipeline orchestration, logging and monitoring, data modeling, and release pipelines. The course combines theoretical knowledge with practical, hands-on exercises to ensure participants can apply what they learn in real-world scenarios.

Course Duration: 40 Hours

Prerequisites:

- Basic knowledge of SQL and database concepts
- Familiarity with Azure basics
- Understanding of data engineering principles

Content Coverage :

Module 1: Introduction to Solution Development

- **Overview of End-to-End Solution Development**
    - Understanding the Lifecycle of Solution Development
    - Key Milestones and Deliverables
- **Importance and Benefits of Integrated Solutions**
    - Streamlining Processes
    - Enhancing Data Accuracy and Consistency
    - Reducing Operational Costs
- **Key Components of a Comprehensive Solution**
    - Data Integration
    - Scalability and Performance
    - Pipeline Orchestration
    - Data Modelling

- Release Management

Module 2: Data Integration

- **Introduction to Data Integration**
  - Purpose and Objectives
  - Common Challenges in Data Integration
- **Configuring Linked Services**
  - Types of Linked Services: Blob Storage, SQL Databases, Data Lakes, etc.
  - Step-by-Step Configuration Guide
- **Securing Data Connections**
  - Introduction to Azure Key Vault
  - Steps to Use Azure Key Vault
  - Best Practices for Securing Data Connections

Module 3: Scalability and Performance

- **Introduction to Scalability and Performance**
  - Definitions and Key Metrics
  - Importance in Big Data Solutions
- **Optimizing Data Ingestion**
  - Using Spark Pools: Overview and Configuration
  - Serverless SQL Pools: When and How to Use Them
  - Case Studies and Examples Based on Data Volume

Module 4: Pipeline Orchestration and Transformation Techniques

- **Introduction to Pipeline Orchestration**
  - Understanding ETL (Extract, Transform, Load) Processes
  - Key Concepts and Terminology
- **Designing ETL Pipelines**
  - Essential Components of ETL Pipelines

    - Best Practices for Efficient Pipeline Design
- **Using Notebooks/Data Flows for Transformations**
    - Overview of Notebooks and Data Flows
    - Practical Examples and Use Cases
- **Implementing Dynamic Control Flows**
    - Introduction to Metadata-Driven Pipelines
    - Steps to Create and Implement Dynamic Control Flows

Module 5: Logging and Monitoring

- **Introduction to Logging and Monitoring**
    - Importance of Logging and Monitoring
    - Key Concepts and Tools
- **Enabling Pipeline Activity Logs**
    - Step-by-Step Guide to Enable Logging
    - Common Log Patterns and What They Mean
- **Configuring Alerts for Monitoring**
    - Setting Up Alerts: Tools and Techniques
    - Monitoring Performance: Key Metrics and Dashboards
    - Troubleshooting Common Issues

Module 6: Data Modelling

- **Introduction to Data Modelling**
    - Importance and Objectives of Data Modelling
    - Key Concepts and Frameworks
- **Implementing Medallion Architecture**
    - Overview of Bronze, Silver, and Gold Layers
    - Step-by-Step Guide to Implement Medallion Architecture
    - Use Cases and Examples
- **Optimizing the Serving Layer for Performance**

- Best Practices for Data Modelling with Serverless SQL Pools and Spark Pools
- Examples of Optimized Serving Layers
- Enabling Access for End Users: Power BI, Data Scientists, etc.

Module 7: Release Pipeline

- **Introduction to Release Pipeline**
  - Key Concepts in Release Management
  - Benefits of Automated Release Pipelines
- **Setting Up CI/CD Pipelines**
  - Introduction to CI/CD Concepts
  - Step-by-Step Guide Using Azure DevOps
  - Examples of Successful CI/CD Implementations
- **Managing Version Control with Synapse Repos**
  - Overview of Synapse Repos
  - Best Practices for Version Control
  - Step-by-Step Guide for Managing Repos
- **Implementing Automated Branching Strategies**
  - Introduction to Branching Strategies
  - Steps to Implement Automated Branching

Module 8: Data Security and Access Control

- **Authentication & Access Control**
  - Overview of Azure Active Directory (Azure AD) and its importance.
  - Setting up Role-Based Access Control (RBAC) for managing permissions.
  - Using Managed Identity for secure connections.
- **Data Protection**
  - Implementing Row-Level Security (RLS) to control access to specific rows.
  - Applying Column-Level Security (CLS) to restrict field-level data access.

- Configuring Dynamic Data Masking (DDM) to safeguard sensitive information dynamically.