DP-601T00: Implementing a Lakehouse with Microsoft Fabric

Course Duration: 8 Hours (1 Day)

Overview

The DP-601T00: Implementing a Lakehouse with Microsoft Fabric course is designed to educate learners on how to build an end-to-end analytics solution using Microsoft's integrated analytics platform, known as Microsoft Fabric. This course covers the essential aspects of working with Fabric, including the creation and management of lake houses, leveraging Apache Spark for data processing, understanding Delta Lake tables for reliable storage, integrating Dataflows Gen2 for data ingestion, and utilizing Data Factory pipelines for orchestrating and automating data workflows. Through the comprehensive modules, participants will gain hands-on experience in implementing a Lakehouse architecture, from data exploration and transformation to visualization, and will learn best practices for team collaboration within the Fabric ecosystem. By completing this course, learners can prepare for Microsoft Fabric certification and acquire the skills necessary to architect and implement robust analytics solutions on the Microsoft Fabric platform, setting a foundation for future innovation and data-driven decision-making.

Audience Profile

DP-601T00: Implementing a Lakehouse with Microsoft Fabric is a comprehensive course designed for professionals aiming to master end-to-end analytics with Microsoft's cutting-edge data solutions. Target Audience for DP-601T00 Course:

- Data Engineers
- Data Scientists
- Data Analysts
- Solutions Architects
- IT Professionals with a focus on data solutions
- Database Administrators
- Software Developers who specialize in data-oriented applications
- Cloud Technology Specialists
- Business Intelligence Professionals
- Enterprise Architects
- Technical Team Leads managing data projects
- Professionals seeking to understand Microsoft's data ecosystem
- Technical Consultants involved in data platform services
- Data Platform Engineers

Course Syllabus

Module 1: Introduction to end-to-end analytics using Microsoft Fabric

Discover how Microsoft Fabric can meet your enterprise's analytics needs in one platform. Learn about Microsoft Fabric, how it works, and identify how you can use it for your analytics needs.

- Introduction to Microsoft Fabric
- Data teams and Fabric
- Enable and use Microsoft Fabric

Module 2: Get started with lake houses in Microsoft Fabric

Lake houses merge data lake storage flexibility with data warehouse analytics. Microsoft Fabric offers a lake house solution for comprehensive analytics on a single SaaS platform.

- What is a Lakehouse
- Work with a Fabric Lakehouse
- Explore, transform and visualize data in the Lakehouse

Module 3: Use Apache Spark in Microsoft Fabric

Apache Spark is a core technology for large-scale data analytics. Microsoft Fabric provides support for Spark clusters, enabling you to analyze and process data in a Lakehouse at scale.

- Prepare to use Apache Spark
- Run Spark in Fabric
- Load data in a Spark Data Frame
- Transform data in a Spark Data Frame
- Partition the output file
- Work with data using Spark SQL
- Query Data using Spark SQL API
- Visualize Data

Module 4: Work with Delta Lake tables in Microsoft Fabric

Tables in a Microsoft Fabric Lake house are based on the Delta Lake storage format commonly used in Apache Spark. By using the enhanced capabilities of delta tables, you can create advanced analytics solutions.

- Understand Delta Lake
- Create delta tables using code in Spark
- Managed vs External Tables

- Work with delta tables in Spark
- Data versioning and Time Travel
- Use delta tables with Streaming data

Module 5: Ingest Data with Dataflows Gen2 in Microsoft Fabric

Data ingestion is crucial in analytics. Microsoft Fabric's Data Factory offers Dataflows (Gen2) for visually creating multi-step data ingestion and transformation using Power Query Online.

- Understand Dataflows (Gen2)
- Dataflow (Gen2) benefits and limitations
- Explore Dataflows (Gen2) in Microsoft Fabric
- Integrate Dataflows (Gen2) and Pipelines in Microsoft Fabric

Module 6: Use Data Factory pipelines in Microsoft Fabric

Microsoft Fabric includes Data Factory capabilities, including the ability to create pipelines that orchestrate data ingestion and transformation tasks.

- Pipelines in Microsoft Fabric
- Common Activities – Copy Data
- Common Activities – pipeline templates
- Run and monitor pipeline