Azure

DP-601T00: Implementing a Lakehouse with Microsoft Fabric

$695.00

- 1 Day

Upcoming Dates

May 12
Oct 06

Course Description

This course is designed to build your foundational skills in data engineering on Microsoft Fabric, focusing on the Lakehouse concept. This course will explore the powerful capabilities of Apache Spark for distributed data processing and the essential techniques for efficient data management, versioning, and reliability by working with Delta Lake tables. This course will also explore data ingestion and orchestration using Dataflows Gen2 and Data Factory pipelines. This course includes a combination of lectures and hands-on exercises that will prepare you to work with lakehouses in Microsoft Fabric.

Course Outline

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.

Lessons

- Explore end-to-end analytics with Microsoft Fabric
- Data teams and Microsoft Fabric
- Enable and use Microsoft Fabric

Exercises:

- None

Get started with lakehouses in Microsoft Fabric

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

Lessons

- Explore the Microsoft Fabric Lakehouse
- Work with Microsoft Fabric Lakehouses
- Explore and transform data in a lakehouse

Exercise:

- Create and ingest data with a Microsoft Fabric Lakehouse

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.

Lessons

- Prepare to use Apache Spark
- Run Spark code
- Work with data in a Spark dataframe
- Work with data using Spark SQL
- Visualize data in a Spark notebook

Exercise

- Analyze data with Apache Spark

Work with Delta Lake tables in Microsoft Fabric

Tables in a Microsoft Fabric lakehouse 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.

Lessons

- Understand Delta Lake
- Create delta tables
- Work with delta tables in Spark
- Use delta tables with streaming data

Exercise

- Use delta tables in Apache Spark

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.

Lessons

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

Exercise

- Create and use a Dataflow (Gen2) in Microsoft Fabric

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.

Lessons

- Understand pipelines
- Use the Copy Data activity
- Use pipeline templates
- Run and monitor pipelines

Exercise

- Ingest data with a pipeline

Audience

The primary audience for this course is data professionals who are familiar with data modeling, extraction, and analytics. It is designed for professionals who are interested in gaining knowledge about Lakehouse architecture, the Microsoft Fabric platform, and how to enable end-to-end analytics using these technologies.

Prerequisites

Before attending this course, students should be familiar with:

- Fundamental data analytics concepts.
- The Microsoft Fabric interface and core concepts.
- Microsoft Fabric lakehouses and Apache Spark.
- Microsoft Fabric and data orchestration.
- Microsoft Fabric lakehouses, Apache Spark, and SparkSQL.

What You Will Learn