Implementing a Lakehouse with Microsoft Fabric

CODICE	DP-601T00
DURATA	1 gg
PREZZO	420,00 €
EXAM	

DESCRIZIONE

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.

TARGET

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.

PREREQUISTI

You should be familiar with basic data concepts and terminology.

CONTENUTI

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.

Learning objectives

In this module, you'll learn how to:

- Describe end-to-end analytics in Microsoft Fabric

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.

Learning objectives

In this module, you'll learn how to:

- Describe core features and capabilities of lakehouses in Microsoft Fabric
- Create a lakehouse
- Ingest data into files and tables in a lakehouse
- Query lakehouse tables with SQL

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.

Learning objectives

In this module, you'll learn how to:

- Configure Spark in a Microsoft Fabric workspace
- Identify suitable scenarios for Spark notebooks and Spark jobs
- Use Spark dataframes to analyze and transform data
- Use Spark SQL to query data in tables and views
- Visualize data in a Spark notebook

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.

Learning objectives

In this module, you'll learn how to:

- Understand Delta Lake and delta tables in Microsoft Fabric
- Create and manage delta tables using Spark
- Use Spark to query and transform data in delta tables
- Use delta tables with Spark structured streaming

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.

Learning objectives

In this module, you'll learn how to:

- Describe Dataflow (Gen2) capabilities in Microsoft Fabric
- Create Dataflow (Gen2) solutions to ingest and transform data
- Include a Dataflow (Gen2) in a pipeline

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.

Learning objectives

In this module, you'll learn how to:

- Describe pipeline capabilities in Microsoft Fabric
- Use the Copy Data activity in a pipeline
- Create pipelines based on predefined templates
- Run and monitor pipelines