Implementing a Lakehouse with Microsoft Fabric (DP-601)

Duration: 1 Day **Course Code: M-DP601**

Overview:

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 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.

Objectives:

- Students will learn,
- Introduction to end-to-end analytics using Microsoft Fabric
- Get started with lakehouses in Microsoft Fabric
- Use Apache Spark in Microsoft Fabric
- Work with Delta Lake tables in Microsoft Fabric
- Ingest Data with Dataflows Gen2 in Microsoft Fabric
- Use Data Factory pipelines in Microsoft Fabric

Prerequisites:

- You should be familiar with basic data concepts and terminology.

Content:

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

- Describe end-to-end analytics in Microsoft Fabric

Module 2: Get started with lakehouses in Microsoft Fabric

- 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

Module 3: Use Apache Spark in Microsoft Fabric

- 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

Module 4: Work with Delta Lake tables in Microsoft Fabric

- 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

Module 5: Ingest Data with Dataflows Gen2 in Microsoft Fabric

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

Module 6: Use Data Factory pipelines in Microsoft Fabric

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

Further Information:

For More information, or to book your course, please call us on 0800/84.009

info@globalknowledge.be

www.globalknowledge.com/en-be/