Curso MAFAB200 - Engenharia de Dados com o Microsoft Fabric

Duração: 35h

Introdução

À medida que as organizações caminham para decisões baseadas em dados, a capacidade de desenhar e gerir uma plataforma analítica ponta a ponta torna-se essencial.

Neste curso vai aprender a construir uma solução analítica completa com o Microsoft Fabric, da ingestão e transformação dos dados até à modelação e visualização para obter conclusões de negócio. Vai trabalhar com componentes-chave como lakehouses, data pipelines e Spark, e perceber como encaixam numa arquitetura de dados moderna.

Através de exercícios práticos, vai ganhar as competências necessárias para desenhar, implementar e operar soluções de dados escaláveis em Fabric.

Público-alvo

Este curso destina-se a engenheiros de dados e profissionais de BI que queiram construir lakehouses e data warehouses com o Microsoft Fabric, aprender a ingerir e limpar os dados e, por fim, ligá-los ao Power BI.

No final do curso

Após a conclusão deste curso, os formandos deverão ser capazes de:
• Compreender a arquitetura do Microsoft Fabric e do OneLake.
• Ingerir e transformar dados com o Data Factory, pipelines dinâmicos e Dataflow Gen2.
• Fazer engenharia de dados com Spark, PySpark e Spark SQL.
• Trabalhar com tabelas Delta e lakehouses.
• Construir um data warehouse e utilizar bases de dados SQL no Fabric.
• Criar relatórios no Fabric e utilizar o Copilot.

Pré-requisitos

Familiaridade com conceitos de dados e de BI.

Exames de certificação

Em conjunto com o MAFAB300, contribui para a preparação do exame DP-700: Implementing Data Engineering Solutions Using Microsoft Fabric, da certificação Microsoft Certified: Fabric Data Engineer Associate (exame não incluído neste curso).

Conteúdo programático em detalhe

Introdução ao Microsoft Fabric
• O que é o Microsoft Fabric?
• Do data warehouse tradicional ao data lake
• Arquitetura data mesh

- Trabalhar com o Task Flow
- Licenciamento do Microsoft Fabric
- Monitorizar o Microsoft Fabric
- LAB: primeiros passos com o Microsoft Fabric

Introdução aos data lakes

- Introdução aos data lakes
- Trabalhar com o Azure Storage
- Guardar dados num data lake
- A arquitetura medalhão
- Formatos de armazenamento em data lakes: CSV e Parquet
- Delta Lake
- Outros open table formats

Microsoft OneLake

- O que é o OneLake?
- Criar workspaces
- Segurança dos workspaces
- Trabalhar com domains
- Workspaces e controlo de versões: integração com Azure DevOps e GitHub
- Deployment pipelines

Armazenar dados no OneLake

- Criar uma lakehouse
- Carregar dados manualmente na lakehouse
- O SQL analytics endpoint da lakehouse
- Criar um modelo semântico
- Trabalhar com shortcuts
- Shortcuts e segurança
- Ligar aplicações externas ao Microsoft OneLake
- Segurança da lakehouse
- LAB: montar lakehouses no OneLake

Utilizar o Copilot no Microsoft Fabric

- Como o Copilot encaixa no Microsoft Fabric
- Licenciamento e pré-requisitos
- Privacidade dos dados e segurança baseada em papéis

Primeiros passos com o Data Factory

- O que é o Data Factory?
- Criar data pipelines
- A atividade Copy Data
- Trabalhar com o Copy Job
- Executar e monitorizar data pipelines
- LAB: ingerir dados com pipelines

Construir pipelines dinâmicos

- Trabalhar com expressões
- Reutilizar o resultado de uma atividade
- Variáveis e parâmetros

- Usar ciclos e lógica condicional em pipelines
- Depurar uma pipeline
- LAB: criar e depurar pipelines avançadas

Ingerir e transformar dados com Dataflow Gen2

- O que é o Dataflow Gen2
- Criar queries para carregar dados
- Compreender a interface do Power Query
- Aplicar transformações
- Query folding
- Controlar o destino da tabela
- Usar dataflows dentro de uma pipeline
- Gerir ligações
- Controlar o staging
- LAB: ingerir e transformar dados com dataflows

Engenharia de dados com Spark

- Introdução ao Apache Spark
- Criar ambientes para clusters de Apache Spark
- Trabalhar com notebooks no Fabric
- Magic commands
- DataFrames do Spark
- Agendar notebooks
- Guia de decisão do Microsoft Fabric: atividade Copy, Dataflow ou Spark
- Usar notebooks de Python
- LAB: primeiros passos com notebooks no Microsoft Fabric

Data wrangling com PySpark e Spark SQL

- Limpeza de dados com PySpark
- Agrupar e agregar dados em PySpark
- Juntar DataFrames
- Usar o Spark SQL para selecionar e manipular dados
- Visualizar dados com notebooks e DataFrames
- LAB: preparação de dados com PySpark e Spark SQL

Trabalhar com tabelas Delta

- O que é um Delta Lake?
- Trabalhar com tabelas Delta
- Gerir alterações de schema
- Versionar e otimizar delta tables
- LAB: trabalhar com delta tables

Construir um data warehouse no Fabric

- O SQL analytics endpoint da lakehouse
- Criar tabelas num Synapse Data Warehouse
- Ingerir dados com pipelines
- Ingerir dados com T-SQL
- Consultar o warehouse
- O modelo semântico de Power BI por omissão

• LAB: criar e usar um warehouse

Bases de dados SQL no Fabric

• O que é a Fabric SQL Database
• Ligar clientes à base de dados
• Controlar a segurança
• Recuperação de desastre
• Fabric SQL Database vs. Fabric Warehouse

Reporting no Fabric

• Criar relatórios de Power BI
• DirectQuery vs. Import com o Microsoft OneLake
• Usar e configurar o modo Direct Lake
• LAB: criar relatórios de Power BI