Curso MAAZDBX - Engenharia de Dados com o Azure Databricks

Duração: 35h

Introdução

Neste curso vai aprender a usar o Azure Databricks para desenhar e implementar soluções de dados ponta a ponta. Vai ganhar experiência prática na criação de workspaces, no processamento e transformação de dados e na construção de data pipelines escaláveis com Spark, Delta Lake e as ferramentas Databricks.

No fim da formação será capaz de desenhar e operar uma arquitetura lakehouse, permitindo engenharia de dados, analítica e reporting eficientes em Databricks.

Público-alvo

Este curso destina-se a engenheiros de dados e profissionais de analítica que pretendam construir soluções de dados com o Azure Databricks.

No final do curso

Após a conclusão deste curso, os formandos deverão ser capazes de:
• Configurar o Azure Databricks, a computação e o armazenamento em data lakes.
• Governar os dados com o Unity Catalog.
• Trabalhar com notebooks e aceder a dados.
• Construir um lakehouse e fazer data warehousing com Databricks SQL.
• Orquestrar a ingestão e o processamento com Lakeflow Connect e Lakeflow Jobs.
• Ligar o Databricks ao Power BI.

Pré-requisitos

Não são necessários conhecimentos prévios de Azure Databricks.

Exames de certificação

Contribui para a preparação do exame DP-750: Implementing Data Engineering Solutions Using Azure Databricks, da certificação Microsoft Certified: Azure Databricks Data Engineer Associate (exame não incluído neste curso).

Conteúdo programático em detalhe

Primeiros passos com o Azure Databricks
• O que é o Azure Databricks
• Introdução ao Apache Spark
• Workspaces no Azure Databricks
• Aprovisionar workspaces do Azure Databricks
• Navegar nos workspaces
• Configuração e segurança do Azure Databricks
• Databricks Genie One
• Preçário do Azure Databricks

- Funcionalidades de assistência por IA do Databricks
- Databricks Free Edition
- LAB: primeiros passos com o Azure Databricks

Azure Storage e data lakes

- Guardar dados no Azure Databricks
- Uma introdução ao Azure Storage
- Aceder a uma Azure Storage Account
- Guardar dados num data lake
- A arquitetura medalhão
- Formatos de armazenamento em data lakes
- Delta Lake
- Outros open table formats
- LAB: aprovisionar uma Azure Storage Account

Primeiros passos com o Unity Catalog

- Introdução ao Unity Catalog
- Criar um metastore do Unity Catalog
- Configurar o armazenamento do metastore do Unity Catalog
- Navegar no Unity Catalog
- Trabalhar com schemas, tables e volumes
- LAB: montar e configurar um metastore do Unity Catalog

Introdução ao modelo de objetos do Unity Catalog

- Introdução ao modelo de objetos do Unity Catalog
- Trabalhar com catalogs
- Trabalhar com schemas
- Criar tables, views e mais
- O que são os volumes?
- Criar managed volumes
- Criar external volumes
- Tags no Unity Catalog
- Guardar segredos no Unity Catalog
- O futuro do Unity Catalog
- LAB: montar e configurar um metastore do Unity Catalog

Configurar a computação em Databricks

- Apache Spark
- O Databricks Runtime
- Tipos de compute do Databricks
- Tipos de provisioned compute
- Databricks serverless compute
- Ligar notebooks ao compute
- Monitorização da utilização
- LAB: criar e usar compute do Databricks

Utilizar notebooks no Azure Databricks

- O Databricks File System (DBFS)
- Trabalhar com notebooks no Databricks

- Magic commands
- Databricks Utilities
- Genie Code
- Monitorizar a faturação do Genie Code
- Trabalhar com widgets de IPython
- Trabalhar com widgets do Databricks
- Dashboards de notebooks
- Agendar notebooks
- LAB: usar notebooks no Azure Databricks

Aceder a dados no Azure Databricks

- A framework Spark
- Introdução aos DataFrames do Spark
- Ler e escrever dados com DataFrames do Spark
- Montar armazenamento Azure Blob e Data Lake Gen2
- Limpar e transformar dados com a API de DataFrames do Spark
- Schemas e tables no Databricks
- Managed tables vs. unmanaged tables
- Tables no Unity Catalog
- LAB: trabalhar com dados no Azure Databricks

Construir um lakehouse com o Azure Databricks

- Implementar um Delta Lake
- Trabalhar com tabelas Delta
- Gerir alterações de schema
- Versionar e otimizar delta tables
- Data skipping e Z-order
- Delta tables e change data feeds
- Delta tables e o Unity Catalog
- Proteger tables no Unity Catalog
- LAB: construir uma lakehouse com delta tables

Data warehousing e análise com Databricks SQL

- O que são os SQL warehouses?
- Tipos de compute dos SQL warehouses
- Escrever queries com o SQL Editor
- Trabalhar com parâmetros de query
- Acrescentar visualizações a uma query
- Criar e usar dashboards de AI/BI
- Usar o Genie Code na criação de dashboards
- Usar os alertas do Databricks SQL
- Criar agentes Genie
- Monitorizar o Databricks SQL
- LAB: usar SQL warehouses

Introdução ao Databricks Lakeflow

- Introdução ao Databricks Lakeflow
- Pipelines procedimentais vs. declarativas

- Panorâmica do Lakeflow Connect
- Panorâmica das Lakeflow Declarative Pipelines
- Panorâmica dos Lakeflow Jobs

Lakeflow Connect

- O panorama da ingestão de dados no Databricks
- O que são os managed connectors
- Ingerir dados de SQL Server
- O que são os standard connectors
- Ingerir dados de cloud object storage com o Auto Loader
- Faturação do Lakeflow Connect
- LAB: ingerir dados com o Lakeflow Connect

Lakeflow Jobs

- O que são os Lakeflow Jobs?
- Criar Lakeflow Jobs
- Panorâmica das job tasks
- Definir restrições de precedência entre tasks
- Trabalhar com parâmetros e valores de task
- Condições e ciclos em Lakeflow Jobs
- Trabalhar com job triggers
- Monitorizar Lakeflow Jobs
- LAB: criar e executar Lakeflow Jobs

Databricks e Power BI

- Introdução ao Power BI
- Ligar o Power BI Desktop ao Databricks com o Partner Connect
- Ligar o Power BI Desktop ao Databricks manualmente
- LAB: ligar o Power BI ao Databricks