

Pokročilý Data Engineering v Microsoft Fabric

Kód kurzu: GOC681

Pokročilé školení pro datové profesionály, kteří chtějí ovládnout moderní data engineering v Microsoft Fabric s důrazem na praktickou práci v Pythonu a PySparku. Většinu času strávíte psaním kódu v Notebooks - budete implementovat datové transformace pomocí Pythonu (Polars, DuckDB) nebo PySparku, automatizovat ETL procesy a pracovat s pokročilými technikami zpracování dat v distribuovaném prostředí. Naučíte se navrhovat a implementovat medaillon architekturu v prostředí Lakehouse. Prozkoumáte různé metody získávání dat - od Dataflows Gen2 přes orchestrační Pipelines až po vlastní kód v Notebooks. Zvládnete ukládání dat, pochopíte rozdíly mezi datovými sklady a Lakehouses, jejich dotazování a pokročilé komponenty jako stored procedures, functions a data masking. Automatizace a orchestrace datových workflow pomocí Pipelines vás naučí koordinovat komplexní procesy a integrovat jednotlivé vrstvy medaillon architektury. Zaměříte se na optimalizaci výkonu - partitioning, kompresi dat a optimalizaci Spark jobů. Naučíte se monitorovat kapacity Fabric a měřit efektivitu zpracování. Prakticky si vyzkoušíte verzování kódu a nasazování změn pomocí Git integrace a deployment pipelines. Toto školení spolu se školením Pokročilé techniky datové analýzy a reportingu v Microsoft Fabric [GOC682] tvoří přípravu na certifikační zkoušku DP-600: Fabric Analytics Engineer Associate.

Co Vás naučíme

- Navrhovat a implementovat medaillon architekturu v Microsoft Fabric v prostředí Lakehouse
- Implementovat datovou logiku a transformace pomocí Pythonu (Polars, DuckDB) a PySparku v Notebooks
- Pracovat s různými metodami načítání dat - Dataflows Gen2, Pipelines a vlastní kód
- Kopírovat a znovu využívat data v rámci OneLake
- Profilovat, čistit a transformovat data pomocí kódu v různých praktických scénářích
- Pracovat s Lakehouse a Data Warehouse včetně zabezpečení dat
- Automatizovat a orchestrovat datová workflow pomocí Pipelines
- Optimalizovat výkon (partitioning, komprese, optimalizace Spark jobů)
- Verzovat kód a nasazovat změny pomocí Git integrace a deployment pipelines

Pro koho je kurz určen

Kurz je určen především pro datové inženýry a vývojáře, kteří chtějí pracovat s Microsoft Fabric na úrovni kódu a navrhovat, implementovat a provozovat datová řešení v produkčním prostředí. Kurz je vhodný také pro pokročilé analytiky a datové architekty, kteří mají zkušenost s Pythonem a chtějí se posunout směrem k data engineeringu a práci v distribuovaném prostředí.

Požadované vstupní znalosti

- Základní znalost prostředí Microsoft Fabric alespoň v rozsahu kurzu GOC680
- Znalost Pythonu (pandas, list comprehensions, funkce, error handling) a PySparku alespoň v rozsahu kurzu GOC685
- Základní znalost relačních databází a jazyka SQL
- Základní zkušenost s data warehouses nebo data lakes
- Porozumění konceptům extrakce, načítání, profilování a transformace dat
- Zkušenost s nástroji pro datovou analýzu a integraci dat (ETL procesy, data pipelines)
- Znalost verzování a Git integrace výhodou

Osnova kurzu

1. Nastavení prostředí a základní principy
 - Medaillon architektura - principy a komponenty
 - Lakehouse, Data Warehouse, analytické enginy, sémantické vrstvy
 - Nastavení tenantu, volba kapacity, dopady na výkon a cenu
2. Načítání a kopírování dat

GOPAS Praha
Na Strži 2097/63
140 00 Praha 4 - Krč
Tel.: +420 226 201 390
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 530 513 590
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 902 903 132
info@gopas.sk


Copyright © 2026 GOPAS, a.s.,
All rights reserved

Pokročilý Data Engineering v Microsoft Fabric

- Metody načítání dat
- Dataflows Gen2
- Pipelines
- Vlastní ingest pomocí Python / PySpark v Notebooks
- Kopírování a znovupoužití dat v OneLake
- Shortcuts
- Rozhodovací metodika a architektonické dopady
- Praktická implementace

3. Data profiling, čištění a transformace

- Data profiling
- Principy a metody
- Implementace v Pythonu / PySparku (Notebooks)
- Čištění a transformace dat
- Návrh čistících mechanismů na základě profilace
- Transformace dat pomocí kódu
- Slowly changing dimensions a pokročilé scénáře

4. Uložení dat

- Lakehouse vs. Data Warehouse – rozdíly a použití
- Dotazování nad daty
- SQL dotazy
- Dotazy nad Lakehouse a Warehouse
- Pokročilé komponenty
- Stored procedures, functions, role, schémata
- RLS, CLS, data masking

5. Automatizace

- Orchestrační Pipelines
- Koordinace a závislosti
- Integrace notebooků, dataflows a SQL objektů
- Orchestrace notebooků
- Řízení návazných kroků v Python / PySpark
- Fail-over a error handling

6. Monitoring a optimalizace

- Optimalizace výkonu Spark workloadů
- Partitioning, komprese, V-order, vacuuming
- Monitoring Fabric kapacit a efektivitu zpracování

7. Verzování a nasazování změn

- Git integrace
- Deployment pipelines

GOPAS Praha

Na Strži 2097/63
140 00 Praha 4 - Krč
Tel.: +420 226 201 390
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 530 513 590
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 902 903 132
info@gopas.sk



Copyright © 2026 GOPAS, a.s.,
All rights reserved