

# Pokročilý Data Engineering v Microsoft Fabric

Kód kurzu: GOC681

Pokročilé školenie určené pre dátových profesionálov, ktorí chcú zvládnuť moderný data engineering v Microsoft Fabric so zameraním na praktickú prácu v Pythone a PySparku. Väčšinu času strávite písaním kódu v Notebooks - budete implementovať dátové transformácie pomocou Pythonu (Polars, DuckDB) alebo PySparku, automatizovať ETL procesy a pracovať s pokročilými technikami spracovania dát v distribuovanom prostredí. Naučíte sa navrhovať a implementovať medallion architektúru v Lakehouse prostredí. Preskúmate rôzne metódy získavania dát - od Dataflows Gen2 cez orchestračné Pipelines až po vlastný kód v Notebooks. Osvojíte si ukladanie dát, pochopíte rozdiely medzi dátovými skladmi a Lakehouses, ich dotazovanie a pokročilé komponenty, ako sú stored procedures, functions či data masking. Automatizácia a orchestrácia dátových workflow pomocou Pipelines vás naučí koordinovať komplexné procesy a integrovať jednotlivé vrstvy medallion architektúry. Zameriate sa na optimalizáciu výkonu - partitioning, kompresiu dát a optimalizáciu Spark jobov. Naučíte sa monitorovať Fabric kapacity a vyhodnocovať efektivitu spracovania. Prakticky si vyskúšate verzovanie kódu a nasadzovanie zmien pomocou Git integrácie a deployment pipelines. Toto školenie spolu so školením Pokročilé techniky dátovej analýzy a reportingu v Microsoft Fabric [GOC682] tvorí kompletnú prípravu na certifikačnú skúšku DP-600: Fabric Analytics Engineer Associate.

## Čo sa naučíte

- Navrhovať a implementovať medallion architektúru v Microsoft Fabric v prostredí Lakehouse
- Implementovať dátovú logiku a transformácie pomocou Pythonu (Polars, DuckDB) a PySparku v Notebooks
- Pracovať s rôznymi metódami načítavania dát - Dataflows Gen2, Pipelines a vlastný kód
- Kopírovať a opätovne využívať dáta v rámci OneLake
- Profilovať, čistiť a transformovať dáta pomocou kódu v rôznych praktických scenároch
- Pracovať s Lakehouse a Data Warehouse vrátane zabezpečenia dát
- Automatizovať a orchestrovať dátové workflow pomocou Pipelines
- Optimalizovať výkon (partitioning, kompresia, optimalizácia Spark jobov)
- Verzovať kód a nasadzovať zmeny pomocou Git integrácie a deployment pipelines

## Pre koho je kurz určený

Kurz je určený predovšetkým pre dátových inžinierov a vývojárov, ktorí chcú pracovať s Microsoft Fabric na úrovni kódu a navrhovať, implementovať a prevádzkovať dátové riešenia v produkčnom prostredí. Kurz je vhodný aj pre pokročilých analytikov a dátových architektov, ktorí majú skúsenosť s Pythonom a chcú sa posunúť smerom k data engineeringu a práci v distribuovanom prostredí.

## Požadované vstupné znalosti

- Základná znalosť prostredia Microsoft Fabric aspoň v rozsahu kurzu GOC680
- Znalosť Pythonu (pandas, list comprehensions, funkcie, error handling) a PySparku aspoň v rozsahu kurzu GOC685
- Základná znalosť relačných databáz a jazyka SQL
- Základná skúsenosť s data warehouses alebo data lakes
- Pochopenie konceptov extrakcie, načítavania, profilovania a transformácie dát
- Skúsenosť s nástrojmi na dátovú analýzu a integráciu dát (ETL procesy, data pipelines)
- Znalosť verzovania a Git integrácie je výhodou

## Osnova kurzu

1. Nastavenie prostredia a základné princípy
  - Medallion architektúra - princípy a komponenty
  - Lakehouse, Data Warehouse, analytické enginy, sémantické vrstvy
  - Nastavenie tenantu, voľba kapacity, dopady na výkon a cenu
2. Načítavanie a kopírovanie dát

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

**GOPAS Brno**  
Nové sady 996/25  
602 00 Brno  
Tel.: +420 542 422 111  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Bratislava**  
Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)

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

# Pokročilý Data Engineering v Microsoft Fabric

- Metódy načítavania dát
  - Dataflows Gen2
  - Pipelines
  - Vlastný ingest pomocou Python / PySpark v Notebooks
  - Kopírovanie a opätovné použitie dát v OneLake
  - Shortcuts
  - Rozhodovacia metodika a architektonické dopady
  - Praktická implementácia
3. Data profiling, čistenie a transformácia
- Data profiling
  - Princípy a metódy
  - Implementácia v Pythone / PySparku (Notebooks)
  - Čistenie a transformácia dát
  - Návrh čistiacich mechanizmov na základe profilovania
  - Transformácia dát pomocou kódu
  - Slowly Changing Dimensions a pokročilé scenáre
4. Ukladanie dát
- Lakehouse vs. Data Warehouse – rozdiely a použitie
  - Dotazovanie nad dátami
  - SQL dotazy
  - Dotazovanie nad Lakehouse a Warehouse
  - Pokročilé komponenty
  - Stored procedures, functions, role, schémy
  - RLS, CLS, data masking
5. Automatizácia
- Orchestračné Pipelines
  - Koordinácia a závislosti
  - Integrácia notebookov, dataflows a SQL objektov
  - Orchestrácia notebookov
  - Riadenie nadväzujúcich krokov v Pythone / PySparku
  - Fail-over a error handling
6. Monitoring a optimalizácia
- Optimalizácia výkonu Spark workloadov
  - Partitioning, kompresia, V-order, vacuuming
  - Monitoring Fabric kapacít a efektivity spracovania
7. Verzovanie a nasadzovanie zmien
- Git integrácia
  - Deployment pipelines

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

**GOPAS Brno**  
Nové sady 996/25  
602 00 Brno  
Tel.: +420 542 422 111  
[info@gopas.cz](mailto:info@gopas.cz)

**GOPAS Bratislava**  
Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 902 903 132  
[info@gopas.sk](mailto:info@gopas.sk)



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