

Data Engineering v Microsoft Fabric

Kód kurzu: GOC681

Školení je ideální pro datové profesionály, kteří chtějí posunout své dovednosti a plně ovládnout Data Engineering v prostředí Microsoft Fabric. Toto školení spolu se školením Data Analysis v Microsoft Fabric [GOC682] tvoří dohromady důkladnou přípravu na certifikační zkoušku DP-600: Fabric Analytics Engineer Associate. Naučíte se principy medallion architektury, prozkoumáte metody získávání dat pomocí Dataflows Gen2, Pipelines a Notebooks. Kurz pokrývá ukládání dat, rozdíly mezi datovými sklady a Lakehouses, jejich dotazování a komponenty jako stored procedures, functions, data masking, atd. Automatizace procesů pomocí orchestračních pipelines vás naučí koordinovat data workflows a integrovat s medallion architekturou. Z pohledu optimalizace výkonu se zaměříte na techniky jako partitioning a komprese. Naučíte se monitorovat kapacity a měřit efektivitu zpracování dat. Zkusíte si verzování a nasazování změn pomocí Git integrace a deployment pipelines.

| Pobočka | Dnů | Cena kurzu | ITB |
|------------|-----|------------|-----|
| Praha | 3 | 29 600 Kč | 30 |
| Brno | 3 | 29 600 Kč | 30 |
| Bratislava | 3 | 1 250 € | 30 |

Uvedené ceny jsou bez DPH.

Termíny kurzu

| Datum | Dnů | Cena kurzu | Typ výuky | Jazyk výuky | Lokalita |
|-------|-----|------------|-----------|-------------|----------|
|-------|-----|------------|-----------|-------------|----------|

Uvedené ceny jsou bez DPH.

Co Vás naučíme

- Porozumíte principům a komponentům medallion architektury
- Naučíte se, jak efektivně nastavit prostředí a konfigurovat nastavení tenantu
- Prozkoumáte různé metody extrakce a načítání dat pomocí Dataflows Gen2, Pipelines a Notebooks
- Zvládnete techniky kopírování a opětovného použití dat v OneLake
- Ovládnete profilování, čištění a transformaci dat pomocí praktických příkladů
- Ponoříte se do možností ukládání dat, včetně Data Warehouse a Data Lakes, a naučíte se, jak zabezpečit data
- Objevíte, jak automatizovat Dataflows pomocí orchestračních pipelines
- Naučíte se, jak monitorovat a optimalizovat výkon pomocí pokročilých technik
- Pochopíte verzování a nasazování změn pomocí Git integrace a deployment pipelines

Pro koho je kurz určen

Kurz je zaměřen na datové inženýry a vývojáře, kteří se chtějí naučit efektivně pracovat v prostředí Microsoft Fabric. Dále je kurz vhodný pro business analytiky a datové architektky, kteří si chtějí zlepšit své dovednosti v oblasti správy a optimalizace dat.

Požadované vstupní znalosti

- Základní znalost relačních databází a jazyka SQL
- Základní zkušenost s data warehouses nebo data lakes
- Základní porozumění konceptům extrakce, načítání, profilování a transformace dat
- Základní zkušenost s nástroji pro datovou analýzu a integraci dat (např. ETL procesy, data pipelines)
- Znalost verzování a Git integrace výhodou

Osnova kurzu

1. Nastavení prostředí a vysvětlení principů

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved

Data Engineering v Microsoft Fabric

- Medallion architecture - principy a komponenty:
 - Data Lakes, Data Warehouses, strukturovaná analytika, sémantické modely, analytické engines
 - nastavení tenantu, jak vybrat kapacitu, jak nad tím uvažovat
2. Data Ingestion a kopírování dat
- Průzkum způsobů načtení dat
 - Dataflows Gen2
 - Pipelines
 - Notebooks
 - Kopírování a přepoužití dat v rámci OneLake
 - shortcuts
 - rozhodovací metodika přístupu
 - kdy jakou metodu zvolit, jak nad tím architektonicky přemýšlet + praktická aplikace, výroba
3. Data profiling, čištění a transformace
- Data profiling
 - principy
 - implementace a možnosti v notebooks a dataflows gen2
 - Data cleaning and transforming
 - sestavení čistících mechanismů na základě výsledků datové profilace
 - implementace datových transformací
 - slowly changing dimensions, atd.
4. Uložení dat
- Rozdíly mezi Data Warehousem a Lakehousem a jejich vytvoření
 - Dotazování na data v rámci DWH a Lakehouse
 - SQL dotazy a visual queries
 - vnitroitemové dotazy, meziitemové dotazy
 - Komponenty v rámci DWH a Lakehouse
 - store procedures, functions, roles, schemas, RLS, CLS, data masking, atd.
 - kdy použít warehouse, lakehouse, jak pracovat s jejich daty, jak vytvářet jejich podružné komponenty, jak zabezpečit data
5. Automatizace
- Orchestrační pipeline
 - Koordinace – sekvencování data workflows
 - Závislosti - správa pořadí exekucí
 - Integrace - propojení s komponenty medallion architektury
 - Vnitřní orchestrace notebooků
 - jak využít data pipelines pro orchestraci návazných pipelines, notebooků, dataflows, store procedures
 - jak orchestrovat notebooky v rámci aktivní session
 - implementace fail-over scénářů
6. Monitoring & optimalizace
- Optimalizace výkonu
 - Bližší pohled na techniky vylepšující výkon
 - Partitioning
 - komprese
 - V-order
 - vacuuming
 - Admin perspektiva

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved

Data Engineering v Microsoft Fabric

- monitoring kapacit – sledování využití zdrojů
- metriky efektivity – měření efektivity zpracování dat

7. Verzování a nasazování změn

- Git integrace
- Deployment pipelines

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved