

Data Engineering v Microsoft Fabric

Kód kurzu: GOC681

Školenie je ideálne pre dátových profesionálov, ktorí chcú posunúť svoje zručnosti a plne ovládnuť Data Engineering v prostredí Microsoft Fabric. Toto školenie spolu so školením Data Analysis v Microsoft Fabric [GOC682] tvoria dohromady dôkladnú prípravu na certifikačnú skúšku DP-600: Fabric Analytics Engineer Associate. Naučíte sa princípy medaillon architektúry, preskúmate metódy získavania dát pomocou Dataflows Gen2, Pipelines a Notebooks. Kurz pokrýva ukladanie dát, rozdiely medzi dátovými skladmi a Lakehouses, ich dotazovanie a komponenty ako stored procedures, functions, data masking atď. Automatizácia procesov pomocou orchestračných pipelines vás naučí koordinovať data workflows a integrovať s medaillon architektúrou. Z pohľadu optimalizácie výkonu sa zameriate na techniky ako partitioning a kompresia. Naučíte sa monitorovať kapacity a merať efektivitu spracovania dát. Skúsíte si verzovanie a nasadzovanie zmien pomocou Git integrácie a deployment pipelines.

Čo Vás naučíme

- Porozumiete princípom a komponentom medallion architektúry
- Naučíte sa, ako efektívne nastaviť prostredie a konfigurovať nastavenia tenantu
- Preskúmate rôzne metódy extrakcie a načítania dát pomocou Dataflows Gen2, Pipelines a Notebooks
- Zvládnete techniky kopírovania a opätovného použitia dát v OneLake
- Ovládnete profilovanie, čistenie a transformáciu dát pomocou praktických príkladov
- Ponoríte sa do možností ukladania dát, vrátane Data Warehouse a Data Lakes, a naučíte sa, ako zabezpečiť dáta
- Objavíte, ako automatizovať Dataflows pomocou orchestračných pipelines
- Naučíte sa, ako monitorovať a optimalizovať výkon pomocou pokročilých techník
- Pochopíte verzovanie a nasadzovanie zmien pomocou Git integrácie a deployment pipelines

Pre koho je kurz určený

Kurz je zameraný na dátových inžinierov a vývojárov, ktorí sa chcú naučiť efektívne pracovať v prostredí Microsoft Fabric.

Ďalej je kurz vhodný pre business analytikov a dátových architektov, ktorí si chcú zlepšiť svoje zručnosti v oblasti správy a optimalizácie dát.

Požadované vstupné znalosti

- Základná znalosť relačných databáz a jazyka SQL
- Základné skúsenosti s data warehouses alebo data lakes
- Základné porozumenie konceptom extrakcie, načítania, profilovania a transformácie dát
- Základné skúsenosti s nástrojmi pre dátovú analýzu a integráciu dát (napr. ETL procesy, data pipelines)
- Znalosť verzovania a Git integrácie výhodou

Osnova kurzu

1. Nastavenie prostredia a vysvetlenie princípov
 - Medaillon architecture – princípy a komponenty:
 - Data Lakes, Data Warehouses, štruktúrovaná analytika, sémantické modely, analytické engines
 - Nastavenie tenantov, ako vybrať kapacitu, ako nad tým uvažovať
2. Data Ingestion a kopírovanie dát
 - Prieskum spôsobov načítania dát
 - Dataflows Gen2
 - Pipelines
 - Notebooks
 - Kopírovanie a prepoužitie dát v rámci OneLake
 - Shortcuts
 - Rozhodovacia metodika prístupu
 - Kedy, akú metódu zvoliť, ako nad tým architektonicky premýšľať + praktická aplikácia, výroba

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

3. Data profiling, čistenie a transformácia

- Data profiling
- Princípy
- Implementácia a možnosti v notebooks a dataflows gen2
- Data cleaning and transforming
- Zostavenie čistiacich mechanizmov na základe výsledkov dátovej profilácie
- Implementácia dátových transformácií
- Slowly changing dimensions atď.

4. Uloženie dát

- Rozdiely medzi Data Warehouse a Lakehouse a ich vytvorenie
- Dotazovanie na dáta v rámci DWH a Lakehouse
- SQL dotazy a visual queries
- Vnútroitemové dotazy, medziitemové dotazy
- Komponenty v rámci DWH a Lakehouse
- Store procedures, functions, roles, schemas, RLS, CLS, data masking atď.
- Kedy použiť warehouse, lakehouse, ako pracovať s ich dátami, ako vytvárať ich podružné komponenty, ako zabezpečiť dáta

5. Automatizácia

- Orchestračné pipeline
- Koordinácia – sekvencovanie data workflows
- Závislosti – správa poradia exekúcií
- Integrácia – prepojenie s komponentmi medallion architektúry
- Vnútroiná orchestrácia notebookov
- Ako využiť data pipelines pre orchestráciu nadväzných pipelines, notebookov, dataflows, store procedures
- Ako orchestrovať notebooky v rámci aktívnej session
- Implementácia fail-over scenárov

6. Monitoring & optimalizácia

- Optimalizácia výkonu
- Bližší pohľad na techniky vylepšujúce výkon
- Partitioning
- Kompresia
- V-order
- Vacuuming
- Admin perspektíva
- Monitoring kapacít – sledovanie využitia zdrojov
- Metriky efektivity – meranie efektivity spracovania dát

7. Verzovanie a nasadzovanie zmien

- Git integrácia
- 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