

# Azure Data Factory

Kód kurzu: GOC647

Kurz je určený pre dátových inžinierov a tvorcov ETL/ELT procesov v cloudovom prostredí Microsoft Azure, ktorí sa chcú naučiť efektívne implementovať ETL/ELT procesy s využitím služby Azure Data Factory, ktorými budú načítavať dáta z rôznych dátových zdrojov, transformovať ich a vkladať do dátového skladu. V úvode kurzu sa zoznámite so službou Azure Data Factory, jej kľúčovými vlastnosťami a komponentmi a dozviete sa, aké sú odlišnosti implementácie ETL procesov v cloude v porovnaní s on-premises prostredím. Naučíte sa v Azure Data Factory implementovať vlastné komplexné Pipelines, zoznámite sa s rôznymi spôsobmi extrakcie dát zo zdrojových systémov v rôznych formátoch a vyskúšate si extrahované dáta transformovať a zapísať do cieľového úložiska. Ďalej sa tiež naučíte vytvorené Pipelines verzovať, nasadiť s využitím CI/CD, automatizovať a monitorovať ich spúšťanie a vhodne zabezpečiť.

## Čo Vás naučíme

- Zoznámite sa so službou Azure Data Factory a jej kľúčovými vlastnosťami a funkciami
- Pochopíte, ako sa odlišujú ETL/ELT procesy v cloude od tých on-premises
- Dozviete sa, ako implementovať komplexné Pipelines v Azure Data Factory
- Naučíte sa načítavať dáta v rôznych formátoch pomocou Azure Data Factory
- Vyskúšate si transformovať načítané dáta pred ich zápisom do cieľového dátového skladu
- Zistíte, ako správne riadiť vývoj Pipelines a ako ich verzovať a následne nasadzovať pomocou CI/CD
- Zoznámite sa s možnosťami automatizácie spúšťania Pipelines a monitoringom ich behu a prípadným ladením problémov

## Pre koho je kurz určený

Kurz je cieleňý na dátových inžinierov a tvorcov ETL/ELT procesov v cloudovom prostredí Microsoft Azure, ktorí sa chcú naučiť efektívne implementovať ETL/ELT procesy s využitím služby Azure Data Factory.

## Požadované vstupné znalosti

- Znalosti základných princípov návrhu dátových skladov a implementácia ETL procesov na úrovni kurzu GOC645
- Znalosti základnej obsluhy cloudového prostredia na úrovni kurzu MOC AZ-900
- Základná znalosť dátových služieb v Microsoft Azure na úrovni kurzu MOC DP-900

## Osnova kurzu

### 1 Úvod do ETL/ELT v Microsoft Azure

- Základná orientácia v pojmoch a prostredí Microsoft Azure
- Prehľad ETL/ELT technológií v Microsoft Azure
- Prehľad súvisiacich technológií
- Vysvetlenie konceptov Data Warehouse (DWH) v Microsoft Azure

### 2 Komponenty Azure Data Factory

- Úvodné predstavenie Azure Data Factory
- Prehľad Integration Runtimes
- Vysvetlenie a ukážka Linked Services
- Predstavenie a vysvetlenie konceptu Dataset
- Vysvetlenie orchestrácie behu pomocou Pipelines

### 3 Dynamické Azure Data Factory

- Vysvetlenie práce s premennými
- Deklarácia a použitie parametrov
- Práca s výrazmi

### 4 Externé aktivity

- Tvorba uložených procedúr vrátane riadenia transakcií a ošetrovania chýb

#### GOPAS Praha

Kodaňská 1441/46  
101 00 Praha 10  
Tel.: +420 234 064 900-3  
[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.: +421 248 282 701-2  
[info@gopas.sk](mailto:info@gopas.sk)



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

- Volanie uložených procedúr z Azure Data Factory
- Predstavenie Notebookov, ukážka ich tvorby a volania v prostredí Azure Data Factory

## 5 Nasadenie a monitorovanie

- Predstavenie dostupných spôsobov nasadenia
- Vysvetlenie použitia Deployment Templates, ich tvorba a nasadenie
- Predstavenie možností monitoringu a debugingu Pipelines pri ich vývoji
- Možnosti monitorovania už nasadených Pipelines
- Proaktívny monitoring pomocou Alertov v Microsoft Azure