

# Azure Data Factory

Course code: GOC647

Kurz je určen pro datové inženýry a tvůrce ETL/ELT procesů v cloudovém prostředí Microsoft Azure, kteří se chtějí naučit efektivně implementovat ETL/ELT procesy s využitím služby Azure Data Factory, kterými budou načítat data z různých datových zdrojů, transformovat je a ukládat do datového skladu. V úvodu kurzu se seznámíte se službou Azure Data Factory, jejími klíčovými vlastnostmi a komponentami a dozvíte se, jaké jsou odlišnosti implementace ETL procesů v cloudu v porovnání s on-premises prostředím. Naučíte se v Azure Data Factory implementovat vlastní komplexní Pipelines, seznámíte se s různými způsoby extrakce dat ze zdrojových systémů v různých formátech a vyzkoušíte si extrahovaná data transformovat a zapsat do cílového úložiště. Dále se také naučíte vytvořené Pipelines verzovat, nasadit s využitím CI/CD, automatizovat a monitorovat jejich spouštění a vhodně zabezpečit.

## Co Vás naučíme

- Seznámíte se se službou Azure Data Factory a jejími klíčovými vlastnostmi a funkcemi
- Pochopíte, jak se odlišují ETL/ELT procesy v cloudu od těch on-premises
- Dozvíte se, jak implementovat komplexní Pipelines v Azure Data Factory
- Naučíte se načítat data v různých formátech s pomocí Azure Data Factory
- Vyzkoušíte si transformovat načtená data před jejich zápisem do cílového datového skladu
- Zjistíte, jak správně řídit vývoj Pipelines a jak je verzovat a následně nasazovat s pomocí CI/CD
- Seznámíte se s možnostmi automatizace spouštění Pipelines a monitoringem jejich běhu a případným laděním problémů

## Pro koho je kurz určen

Kurz je cílen na datové inženýry a tvůrce ETL/ELT procesů v cloudovém prostředí Microsoft Azure, kteří se chtějí naučit efektivně implementovat ETL/ELT procesy s využitím služby Azure Data Factory.

## Požadované vstupní znalosti

- Znalosti základních principů návrhu datových skladů a implementace ETL procesů na úrovni kurzu GOC645.
- Znalosti základní obsluhy cloudového prostředí na úrovni kurzu MOC AZ-900.
- Základní znalost datových služeb v Microsoft Azure na úrovni kurzu MOC DP-900.

## Osnova kurzu

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

- Základní orientace v pojmech a prostředí Microsoft Azure
- Přehled ETL/ELT technologií v Microsoft Azure
- Přehled souvisejících technologií
- Vysvětlení konceptů Data Warehouse (DWH) v Microsoft Azure

### 2 Komponenty Azure Data Factory

- Úvodní představení Azure Data Factory
- Přehled Integration Runtimes
- Vysvětlení a ukázka Linked Services
- Představení a vysvětlení konceptu Dataset
- Vysvětlení orchestrace běhu s pomocí Pipelines

### 3 Dynamické Azure Data Factory

- Vysvětlení práce s proměnnými
- Deklarace a použití parametrů
- Práce s výrazy

### 4 Externí aktivity

- Tvorba uložených procedur včetně řízení transakcí a ošetření chyb

#### 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

- Volání uložených procedur z Azure Data Factory
- Představení Notebooků, ukázka jejich tvorby a volání v prostředí Azure Data Factory

## 5 Nasazení a monitorování

- Představení dostupných způsobů nasazení
- Vysvětlení použití Deployment Templates, jejich tvorba a nasazení
- Představení možností monitoringu a debugingu Pipelines při jejich vývoji
- Možnosti monitorování již nasazených Pipelines
- Proaktivní monitoring s pomocí Alertů v Microsoft Azure