

Moderní datový sklad a řešení pro analýzu velkých dat v Microsoft Azure

Kód kurzu: MOC DP-203

Cílem kurzu je seznámit účastníky s klíčovými přístupy a technologiemi používanými při tvorbě moderních datových skladů a řešení pro analýzu velkých dat v cloudu Microsoft Azure. V rámci kurzu se seznámíte se službou Azure Synapse Analytics, která v sobě kombinuje relační datový sklad postavený na Microsoft SQL Serveru spolu s moderními přístupy založenými na analýze velkých dat s využitím Apache Spark. Na kurzu se naučíte vybrat a navrhnout vhodné úložiště pro vaše data, dozvíte se, jak implementovat ETL procesy v cloudu a jak následně uložená data dotazovat a analyzovat včetně aplikace strojového učení a prezentace výsledků v Power BI.

Pro koho je kurz určen

- Datovým specialistům, datovým architektům a BI architektům, kteří se chtějí naučit navrhovat řešení pro analýzu dat v cloudu Microsoft Azure
- Datovým analytikům, kteří budou pracovat s analytickými řešeními v Microsoft Azure

Co Vás naučíme

- Zorientujete se v nabídce služeb pro uložení a zpracování velkých objemů dat v Microsoft Azure
- Seznámíte se službami Azure Synapse Analytics, Azure Data Lake Storage Gen2, Azure Stream Analytics a Azure Databricks
- Naučíte se implementovat ETL/ELT procesy v cloudu a načítat data z různých zdrojových systémů do úložiště v cloudu
- Dozvíte se, jak ukládat a spravovat data v big data úložišti a jak navrhovat a používat relační datové sklady v cloudu
- Seznámíte se s možnostmi sběru a zpracování událostí v reálném čase

Požadované vstupní znalosti

- Základní znalosti práce s prostředím Microsoft Azure a správou zdrojů v Microsoft Azure na úrovni kurzu AZ-900
- Základní orientace v možnostech uložení a zpracování dat v Microsoft Azure na úrovni kurzu DP-900

Metody výuky

- Odborný výklad s praktickými ukázkami, cvičení na našich počítačích

Studijní materiály

- Oficiální elektronické studijní materiály společnosti Microsoft v anglickém jazyce.

Osnova kurzu

Modul 1: Úvod do zpracování dat v Microsoft Azure

- Pochopíte základní pojmy související se zpracováním dat v Microsoft Azure
- Seznámíte se službou Azure Synapse Analytics a úložištěm Azure Data Lake Storage

Modul 2: Zpracování dat s využitím Azure Synapse Analytics serverless SQL pools

- Naučíte se dotazovat a parsovat soubory uložené v data lake úložišti s pomocí Azure Synapse Analytics serverless SQL pools
- Dozvíte se, jak používat Azure Synapse Analytics serverless SQL pools pro transformace uložených dat
- Zjistíte, jak vytvářet ucelené databáze nad data lake úložištěm s využitím Azure Synapse Analytics serverless SQL pools

Modul 3: Zpracování dat s pomocí Azure Synapse Apache Spark Pools

- Seznámíte se s technologií Apache Spark a její integrací do služby Azure Synapse Analytics
- Naučíte se analyzovat, transformovat a vizualizovat data s pomocí Apache Spark v prostředí Azure Synapse Analytics
- Dozvíte se, jak používat Delta Lake formát pro tvorbu a uložení tabulek

Modul 4: Práce s datovými sklady v Azure Synapse Analytics

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

Moderní datový sklad a řešení pro analýzu velkých dat v Microsoft Azure

- Seznámíte se s principy návrhu tabulek v relačních datových skladech
- Dozvíte se, jaká jsou specifika implementace tabulek v Azure Synapse Analytics
- Naučíte se efektivně importovat data do tabulek ze souborových formátů s využitím EXTERNAL TABLES či příkazu COPY INTO

Modul 5: Přenosy a transformace dat pomocí Azure Synapse Analytics Pipelines

- Pochopíte principy implementace ETL/ELT řešení pro extrakci, transformaci a načtení dat s využitím Azure Synapse Analytics Pipelines
- Naučíte se implementovat control flows a data flows v Azure Synapse Analytics Pipelines
- Seznámíte se s možnostmi použití Spark Notebooks v rámci Azure Synapse Analytics Pipelines

Modul 6: Hybridní transakční a analytické zpracování dat v Azure Synapse Analytics

- Seznámíte se s principy přístupu Hybrid transactional and analytical processing (HTAP)
- Naučíte se nasadit Azure Synapse Link v kombinaci s Azure Cosmos DB
- Vyzkoušíte si implementovat Azure Synapse Link for SQL

Modul 7: Zpracování streamovaných dat v Azure Stream Analytics

- Naučíte se v reálném čase zpracovávat proud událostí ve službě Azure Stream Analytics
- Dozvíte se, jak zpracované události trvale uložit v rámci služby Azure Synapse Analytics
- Seznámíte se s možnostmi vizualizace real-time událostí ve službě Power BI

Modul 8: Řízení dat napříč organizací

- Pochopíte, proč je důležité řešit centrální katalogizaci a řízení dat ve vaší organizaci
- Dozvíte se, jak využít službu Microsoft Purview pro objevení, katalogizaci a řízení dat ve vaší organizaci
- Naučíte se integrovat Microsoft Purview a Azure Synapse Analytics

Modul 9: Zpracování dat v Azure Databricks

- Seznámíte se s možnostmi služby Azure Databricks
- Naučíte se pracovat s Apache Spark v prostředí Azure Databricks
- Dozvíte se, jak spouštět Azure Databricks notebooks ve službě Azure Data Factory

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