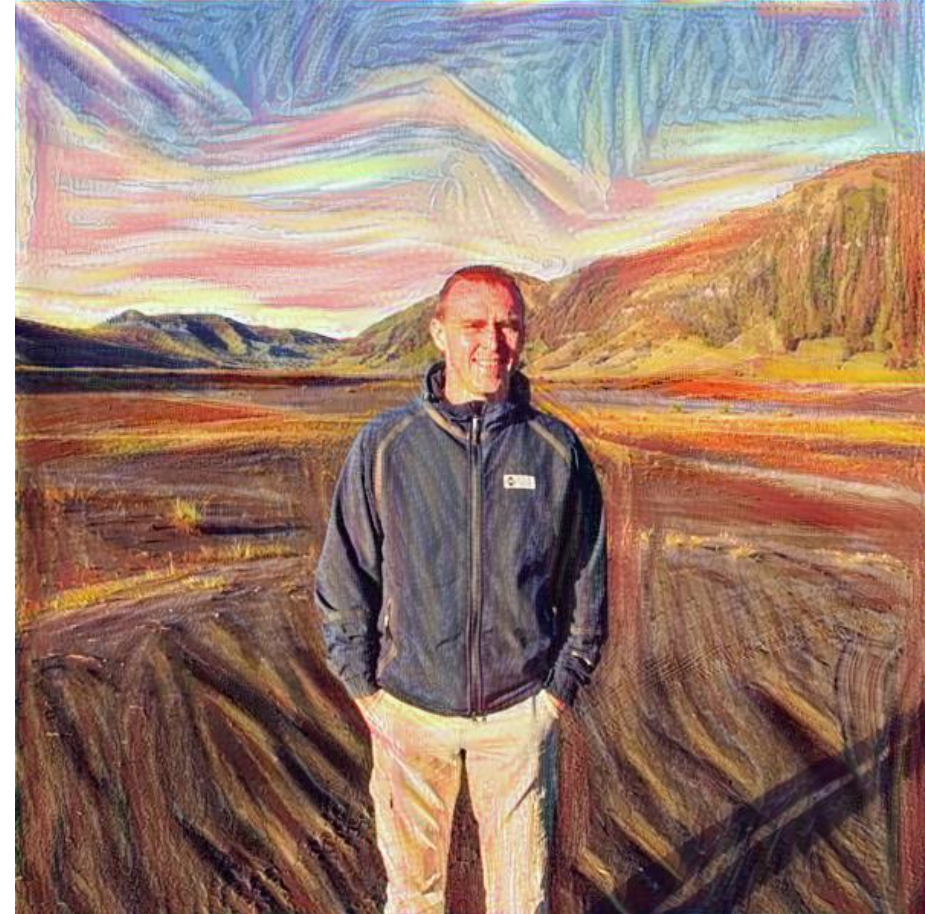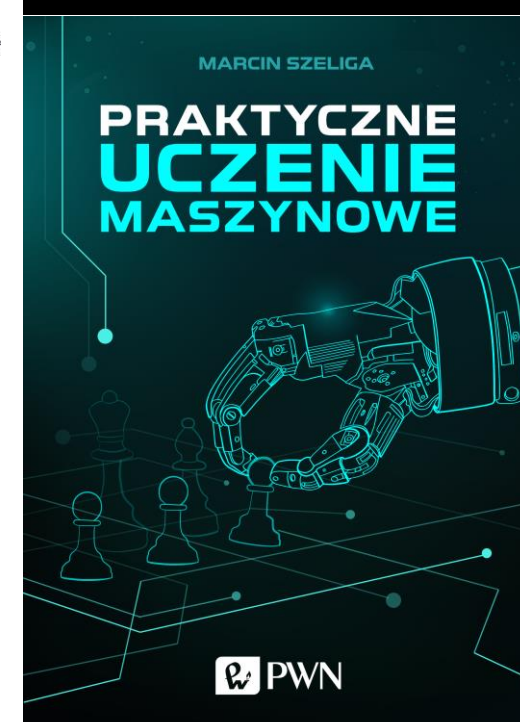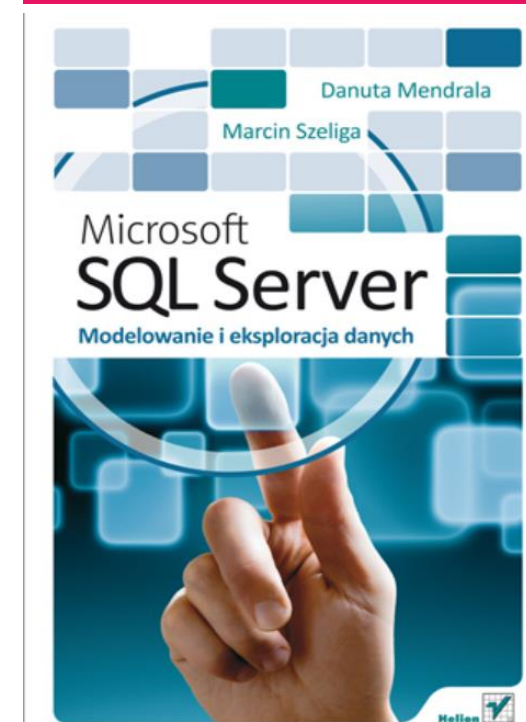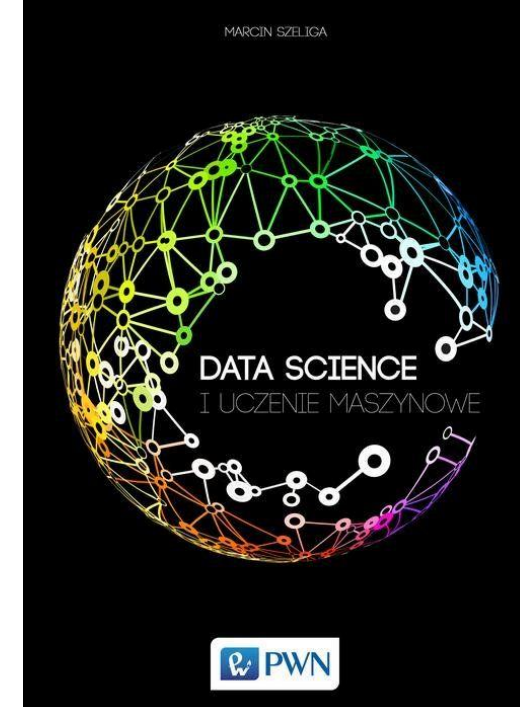# Microsoft & Big Data

Marcin Szeliga

About me

- Machine Learning practitioner
- Over 25 years of professional experience
- Artificial Intelligence MVP & MCT
- Microsoft Certified Solutions Expert
  – Data Management and Analytics
  – Cloud Platform and Infrastructure
  – Business Intelligence
- Microsoft Certified Solutions Developer
  – Azure Solution Architect

Agenda

- Big data fundamentals
  - Big data scenarios
  - Azure
  - Hadoop technology stack
  - Pig and Hive
  - HDInsight
  - Spark
  - Databricks
  - SQL Server 2019 big data clusters
- Understanding Machine Learning

Agenda

- Perform data engineering with Databricks
  - Create a Databricks workspace
  - Apache Spark notebooks
  - Read and write data by using Azure Databricks
  - Perform basic data transformations
  - Perform advanced data transformation
  - Create data pipelines by using Databricks Delta
  - Create data visualizations by using Azure Databricks and Power BI