KOENIG step forward

Microsoft Solutions Partner
Microsoft Cloud Training Services

KOENIG SOLUTIONS · COURSE OUTLINE

Data Transformation Using Spark

Learn to transform data using Apache Spark, Spark SQL and PySpark across Databricks and Azure Synapse Analytics.

10 MODULES

49 TOPICS

COURSE OVERVIEW

This course covers Apache Spark and its architecture within Azure Synapse Analytics and Azure Databricks, followed by Spark SQL and PySpark fundamentals. It then introduces the modern data warehouse and shows how to implement Spark using Databricks and Apache Spark pools. The course continues with reading data from sources such as CSV, JSON, dedicated SQL pools and CosmosDB, transforming that data through notebooks, and writing results to destinations including Azure Data Lake, CosmosDB, dedicated SQL pools and Azure Data Factory. It closes with consuming the transformed data through Power BI.

WHO THIS COURSE IS FOR

Data Engineers | **BI Developers** | **Data Analysts** | **Azure Developers** | **ETL Developers**

LEARNING OBJECTIVES

- Explain Apache Spark architecture and its role in Azure Synapse Analytics
- Describe Spark SQL features, architecture and DataFrames
- Identify PySpark modules, packages and DataFrame concepts
- Describe the architecture and components of a modern data warehouse
- Implement ETL using Azure Databricks and Apache Spark pools
- Apply notebooks to read data from CSV, JSON and CosmosDB
- Configure notebooks to write data to Azure Data Lake and ADF
- Analyse consumption of transformed data through Power BI

COURSE SYLLABUS

10 Modules, 49 Topics

01

Introduction to Spark

TOPICS

Apache Spark overview · What is Apache Spark · Spark pool architecture · Apache Spark in Azure Synapse Analytics · Apache Spark on Azure Databricks

02

Introduction to Spark SQL

TOPICS

Spark SQL - Introduction · Features of Spark SQL · Spark SQL Architecture · Spark SQL - DataFrames · SQLContext

03

Introduction to Spark Python

TOPICS

PySpark - Overview · Who uses PySpark? · Features of PySpark · Advantages of PySpark · PySpark Architecture · PySpark Modules & Packages · PySpark Installation · PySpark DataFrame

04

Introduction to Modern Data Warehouse

TOPICS

Overview of Modern Data Warehouse · Modern Data Warehouse Architecture · Dataflow in Modern Data Warehouse · Components of Modern Data Warehouse · Potential Use Cases

05

Introduction to Databricks / Apache Spark Pool

TOPICS

What is Databricks? · What is Databricks used for? · Common Use Cases for Databricks · Spark Pool Overview · Spark Instances · Quotas and Resource Constraint

06

Implementing Spark with Databricks/Apache Spark Pool

TOPICS

ETL using Azure Databricks · ETL using Apache Spark Pool

07

Reading the Data Using Notebook from Different Sources

TOPICS

Reading data From CSV file · Reading data From JSON file · Reading data From Dedicated SQL Pool · Reading data From CosmosDB

08

Data Transformation Using Databricks

TOPICS

- Creating and using the Notebook in Databricks
- Creating and using the Notebook in Apache Spark Pool
- Using Python in Databricks Notebook
- Using SparkSQL in Databricks Notebook
- Using Python in Apache Spark Pool Notebook
- Using SparkSQL in Apache Spark Pool Notebook

09

Writing the Data Using Notebook to Different Destinations

TOPICS

Writing Data to File in Azure Data Lake · Writing Data to CosmosDB · Writing Data to Dedicated SQL Pool · Sending Data to ADF

10

Consuming Data Using BI Tool

TOPICS

Azure Synapse and PowerBI · Integration of PowerBI in Azure Synapse · PowerBI Service · PowerBI Data Refresh

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training | Live Online Classes (ILO) | Fly-Me-A-Trainer (FMAT) | Flexi (Self-Paced) | 1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com | info@koenig-solutions.com