

# Data Engineering on Microsoft Azure – Intensive Training («DP203»)

Lernen Sie, wie Sie Datentechnik-Workloads in Microsoft Azure implementieren und verwalten, indem sie unter anderem Azure-Dienste wie Azure Synapse Analytics, Azure Data Lake Storage Gen2, Azure Stream Analytics, Azure Databricks verwenden.

**Dauer:** 3 Tage

**Preis:** 2'550.– zzgl. 8.1% MWST

**Herstellercode:** DP-203

## Inhalt

Der Inhalt dieses Intensive Trainings leitet sich aus der Prüfung «[DP-203: Data Engineering on Microsoft Azure](#)» ab. Beginnen Sie schon jetzt auf Microsoft Learn mit der Vorbereitung auf den Kurs. Während der intensiven Trainingstage mit unseren Trainern arbeiten Sie mit den offiziellen Microsoft-Kursunterlagen (mehr Informationen unter «Methodik & Didaktik»).

### Kursinhalt:

#### Modul 1: Erkunden von Compute- und Speicheroptionen für Datentechnikworkloads

Dieses Modul bietet eine Übersicht über die Optionen für Compute- und Speichertechnologien von Azure, die Datentechnikern zur Verfügung stehen, die analytische Workloads erstellen. In diesem Modul werden Methoden zum Strukturieren des Data Lake und zum Optimieren der Dateien für die Untersuchung, das Streaming und die Batchverarbeitung von Workloads vermittelt. Die Kursteilnehmer erfahren, wie sie den Data Lake in Datenoptimierungsebenen organisieren, wenn sie Dateien durch Batch- und Streamverarbeitung transformieren. Anschliessend lernen sie, wie sie Indizes für ihre Datasets erstellen (etwa CSV-, JSON- und Parquet-Dateien) und sie für potenzielle Abfrage- und Workloadbeschleunigung verwenden.

#### Lektionen

- Einführung in Azure Synapse Analytics
- Beschreiben von Azure Databricks
- Einführung in Azure Data Lake Storage
- Beschreiben der Delta Lake-Architektur
- Arbeiten mit Datenströmen mithilfe von Azure Stream Analytics

#### Lab: Erkunden von Compute- und Speicheroptionen für Datentechnikworkloads

- Kombinieren von Streaming und Batchverarbeitung mit einer einzelnen Pipeline
- Organisieren des Data Lake in Ebenen der Dateitransformation
- Indizieren des Data Lake-Speichers für die Beschleunigung von Abfragen und Workloads

#### Modul 2: Ausführen interaktiver Abfragen mithilfe von serverlosen SQL-Pools von Azure Synapse Analytics

In diesem Modul erfahren die Kursteilnehmer, wie sie mit in Data Lake und externen Dateiquellen gespeicherten Dateien arbeiten, indem sie T-SQL-Anweisungen verwenden, die von einem serverlosen SQL-Pool in Azure Synapse Analytics ausgeführt werden. Die Kursteilnehmer fragen Parquet-Dateien ab, die in einem Data Lake gespeichert sind, sowie CSV-Dateien, die in einem externen Datenspeicher gespeichert sind. Als Nächstes erstellen sie Azure Active Directory-Sicherheitsgruppen und erzwingen den Zugriff auf Dateien im Data Lake über rollenbasierte Zugriffssteuerung (Role-Based Access Control, RBAC) und Zugriffssteuerungslisten (Access Control Lists, ACLs).

#### Lektionen

- Kennenlernen von serverlosen SQL-Pool-Funktionen in Azure Synapse
- Abfragen von Daten im Lake mit serverlosen SQL-Pools von Azure Synapse
- Erstellen von Metadatenobjekten in serverlosen SQL-Pools von Azure Synapse
- Schützen von Daten und Verwalten von Benutzern in serverlosen SQL-Pools von Azure Synapse

#### Lab: Ausführen interaktiver Abfragen mithilfe serverloser SQL-Pools

- Abfragen von Parquet-Daten mit serverlosen SQL-Pools
- Erstellen externer Tabellen für Parquet- und CSV-Dateien
- Erstellen von Ansichten mit serverlosen SQL-Pools
- Schützen des Zugriffs auf Daten in einem Data Lake bei Verwendung serverloser SQL-Pools
- Konfigurieren der Data Lake-Sicherheit mit rollenbasierter Zugriffssteuerung (Role-Based Access Control, RBAC) und Zugriffssteuerungslisten (Access Control Lists, ACLs)

### Modul 3: Datenuntersuchung und -transformation in Azure Databricks

In diesem Modul erfahren Sie, wie Sie verschiedene Methoden für Apache Spark-Datenrahmen zum Untersuchen und Transformieren von Daten in Azure Databricks verwenden. Die Kursteilnehmer lernen, wie sie Standardmethoden für Datenrahmen zur Untersuchung und Transformation von Daten ausführen können. Sie lernen auch, wie man erweiterte Aufgaben ausführen, z. B. doppelte Daten entfernen, Datums- / Zeitwerte bearbeiten, Spalten umbenennen und Daten aggregieren kann.

#### Lektionen

- Beschreiben von Azure Databricks
- Lesen und Schreiben von Daten in Azure Databricks
- Arbeiten mit DataFrames in Azure Databricks
- Arbeiten mit erweiterten Methoden für Dataframes in Azure Databricks

#### Lab: Datenuntersuchung und -transformation in Azure Databricks

- Verwenden von Datenrahmen in Azure Databricks zum Untersuchen und Filtern von Daten
- Zwischenspeichern eines Datenrahmens für schnellere nachfolgende Abfragen
- Entfernen doppelt vorhandener Daten
- Bearbeiten von Datums-/Uhrzeitwerten
- Entfernen und Umbenennen von Datenrahmenspalten
- Aggregieren von in einem Datenrahmen gespeicherten Daten

### Modul 4: Untersuchen, Transformieren und Laden von Daten im Data Warehouse mithilfe von Apache Spark

In diesem Modul erfahren Sie, wie Sie in einem Data Lake gespeicherte Daten untersuchen, transformieren und in einen relationalen Datenspeicher laden. Die Kursteilnehmer werden Parquet- und JSON-Dateien untersuchen und Techniken verwenden, um JSON-Dateien mit hierarchischen Strukturen abzufragen und zu transformieren. Anschließend werden die Kursteilnehmer Apache Spark verwenden, um Daten in das Data Warehouse zu laden und Parquet-Daten im Data Lake mit Daten im dedizierten SQL-Pool zu verbinden.

#### Lektionen

- Grundlegendes zu Big-Data-Entwicklung mit Apache Spark in Azure Synapse Analytics
- Erfassen von Daten mit Apache Spark-Notebooks in Azure Synapse Analytics
- Transformieren von Daten mit Dataframes in Apache Spark-Pools in Azure Synapse Analytics
- Integrieren von SQL- und Apache Spark-Pools in Azure Synapse Analytics

#### Lab: Untersuchen, Transformieren und Laden von Daten im Data Warehouse mithilfe von Apache Spark

- Durchführen der Datenuntersuchung in Synapse Studio
- Erfassen von Daten mit Spark-Notebooks in Azure Synapse Analytics
- Transformieren von Daten mit Datenrahmen in Spark-Pools in Azure Synapse Analytics

- Integrieren von SQL- und Spark-Pools in Azure Synapse Analytics

## **Modul 5: Erfassen und Laden von Daten im Data Warehouse**

In diesem Modul lernen die Kursteilnehmer, wie sie Daten mithilfe von T-SQL-Skripts und Synapse Analytics-Integrationspipelines im Data Warehouse erfassen. Die Kursteilnehmer lernen, wie sie Daten mit PolyBase und COPY unter Verwendung von T-SQL in dedizierte Synapse-SQL-Pools laden. Darüber hinaus erfahren die Kursteilnehmer, wie sie die Workloadverwaltung zusammen mit einer Copy-Aktivität in einer Azure Synapse-Pipeline für die Datenerfassung im Petabytebereich verwendet.

### **Lektionen**

- Verwenden von bewährten Methoden zum Laden von Daten in Azure Synapse Analytics
- Datenerfassung im Petabytebereich mit Azure Data Factory

### **Lab: Erfassen und Laden von Daten im Data Warehouse**

- Ausführen der Erfassen im Petabytebereich mit Azure Synapse-Pipelines
- Importieren von Daten mit PolyBase und COPY unter Verwendung von T-SQL
- Verwenden von bewährten Methoden zum Laden von Daten in Azure Synapse Analytics

## **Modul 6: Transformieren von Daten mit Azure Data Factory oder Azure Synapse-Pipelines**

In diesem Modul lernen die Kursteilnehmer, wie sie Datenintegrationspipelines erstellen, um Daten aus mehreren Datenquellen zu erfassen, Daten mithilfe von Zuordnungsdatenflüssen zu transformieren und Daten in eine oder mehrere Datensenken zu verschieben.

### **Lektionen**

- Datenintegration mit Azure Data Factory oder Azure Synapse-Pipelines
- Transformation ohne Code im großen Stil mit Azure Data Factory oder Azure Synapse-Pipelines

### **Lab: Transformieren von Daten mit Azure Data Factory oder Azure Synapse-Pipelines**

- Ausführen von Transformationen ohne Code im grossen Stil mit Azure Synapse-Pipelines
- Erstellen einer Datenpipeline zum Importieren schlecht formatierter CSV-Dateien
- Erstellen von Zuordnungsdatenflüssen

## **Modul 7: Orchestrieren der Datenverschiebung und -transformation in Azure Synapse-Pipelines**

In diesem Modul erfahren Sie, wie Sie verknüpfte Dienste erstellen und die Datenverschiebung und -transformation mithilfe von Notebooks in Azure Synapse-Pipelines orchestrieren.

### **Lektionen**

- Orchestrieren der Datenverschiebung und -transformation in Azure Data Factory

### **Lab: Orchestrieren der Datenverschiebung und -transformation in Azure Synapse-Pipelines**

- Integrieren von Daten aus Notebooks mit Azure Data Factory oder Azure Synapse-Pipelines

## **Modul 8: End-to-End-Sicherheit mit Azure Synapse Analytics**

In diesem Modul erfahren die Kursteilnehmer, wie sie einen Synapse Analytics-Arbeitsbereich und die zugehörige unterstützende Infrastruktur schützen. Die Kursteilnehmer werden den SQL Active Directory-Administrator beobachten, IP-Firewall-Regeln verwalten, Geheimnisse mit Azure Key Vault verwalten und über einen mit Key Vault verknüpften Dienst und Pipelineaktivitäten auf diese Geheimnisse zugreifen. Die Kursteilnehmer lernen, wie sie Sicherheit auf Spaltenebene, Sicherheit auf Zeilenebene und dynamische Datenmaskierung bei Verwendung von dedizierten SQL-Pools implementieren.

### **Lektionen**

- Schützen einer Data Warehouse-Datenbank in Azure Synapse Analytics
- Konfigurieren und Verwalten von Geheimnissen in Azure Key Vault

- Implementieren von Compliancekontrollen für vertrauliche Daten

### Lab: End-to-End-Sicherheit mit Azure Synapse Analytics

- Schützen der unterstützenden Azure Synapse Analytics-Infrastruktur
- Schützen des Azure Synapse Analytics-Arbeitsbereichs und der verwalteten Dienste
- Schützen der Daten im Azure Synapse Analytics-Arbeitsbereich

### Modul 9: Unterstützen von Hybrid Transactional Analytical Processing (HTAP) mit Azure Synapse Link

In diesem Modul erfahren die Kursteilnehmer, wie Azure Synapse Link die nahtlose Konnektivität eines Azure Cosmos DB-Kontos mit einem Synapse-Arbeitsbereich ermöglicht. Die Teilnehmer lernen, wie sie Synapse Link aktivieren und konfigurieren und wie sie anschliessend den Azure-Cosmos-DB-Analysespeicher mithilfe von Apache Spark und serverlosen SQL-Pools abfragen.

#### Lektionen

- Entwerfen der hybriden transaktionalen und analytischen Verarbeitung mithilfe von Azure Synapse Analytics
- Konfigurieren von Azure Synapse Link mit Azure Cosmos DB
- Abfragen von Azure Cosmos DB mit Apache Spark-Pools
- Abfragen von Azure Cosmos DB mit serverlosen SQL-Pools

### Lab: Unterstützen von Hybrid Transactional Analytical Processing (HTAP) mit Azure Synapse Link

- Konfigurieren von Azure Synapse Link mit Azure Cosmos DB
- Abfragen von Azure Cosmos DB mit Apache Spark für Azure Synapse Analytics
- Abfragen von Azure Cosmos DB mit serverlosem SQL-Pool für Azure Synapse Analytics

### Modul 10: Streamverarbeitung in Echtzeit mit Stream Analytics

In diesem Modul erfahren die Kursteilnehmer, wie Streamingdaten mit Azure Stream Analytics verarbeitet werden. Die Kursteilnehmer erfassen Fahrzeugtelemetriedaten in Event Hubs und verarbeiten diese Daten dann in Echtzeit mithilfe verschiedener Fensterfunktionen in Azure Stream Analytics. Die Daten werden in Azure Synapse Analytics ausgegeben. Schliesslich lernen die Kursteilnehmer, wie sie den Stream Analytics-Auftrag skalieren, um den Durchsatz zu erhöhen.

#### Lektionen

- Aktivieren von zuverlässigem Messaging für Big Data-Anwendungen mithilfe von Azure Event Hubs
- Arbeiten mit Datenströmen mithilfe von Azure Stream Analytics
- Erfassen von Datenströmen mit Azure Stream Analytics

### Lab: Streamverarbeitung in Echtzeit mit Stream Analytics

- Verwenden von Stream Analytics zum Verarbeiten von Echtzeitdaten aus Event Hubs
- Verwenden von Stream Analytics-Fensterfunktionen zur Erstellung von Aggregaten und zur Ausgabe in Synapse Analytics
- Skalieren des Azure Stream Analytics-Auftrags, um den Durchsatz durch Partitionierung zu erhöhen
- Neupartitionieren der Streameingabe zur Optimierung der Parallelisierung

### Modul 11: Erstellen einer Streamverarbeitungslösung mit Event Hubs und Azure Databricks

In diesem Modul erfahren die Kursteilnehmer, wie Streamingdaten im grossen Stil mit Event Hubs und Spark Structured Streaming in Azure Databricks erfasst und verarbeitet werden. Die Kursteilnehmer lernen die wichtigsten Funktionen und Einsatzmöglichkeiten von Structured Streaming kennen. Die Teilnehmer implementieren Schiebefenster, um Datenblöcke zu aggregieren und wenden Wasserzeichen an, um veraltete Daten zu entfernen. Schliesslich stellen die Kursteilnehmer eine

Verbindung mit Event Hubs her, um Streams zu lesen und zu schreiben.

## Lektionen

- Verarbeiten von Streamingdaten mit Structured Streaming in Azure Databricks

## Lab: Erstellen einer Streamverarbeitungslösung mit Event Hubs und Azure Databricks

- Erkunden der wichtigsten Features und Verwendungsmöglichkeiten von Structured Streaming
- Streamen von Daten aus einer Datei und Schreiben dieser Daten in ein verteiltes Dateisystem
- Verwenden von gleitenden Fenstern, um Datenblöcke anstelle aller Daten zu aggregieren
- Anwenden von Wasserzeichen zum Entfernen veralteter Daten
- Herstellen einer Verbindung mit Lese- und Schreibstreams für Event Hubs

## Key Learnings

- Erkunden von Rechen- und Speicheroptionen für Data-Engineering-Workloads in Azure
- Entwurf und Implementierung der Serving-Schicht
- Verstehen von Überlegungen zur Datentechnik
- Interaktive Abfragen mit serverlosen SQL-Pools ausführen
- Erforschen, Transformieren und Laden von Daten in das Data Warehouse mit Apache Spark
- Durchführen von Datenexploration und -transformation in Azure Databricks
- Einlesen und Laden von Daten in das Data Warehouse
- Transformieren von Daten mit Azure Data Factory oder Azure Synapse Pipelines
- Integrieren von Daten aus Notebooks mit Azure Data Factory oder Azure Synapse Pipelines
- Optimieren der Abfrageleistung mit Dedicated SQL Pools in Azure Synapse
- Analysieren und Optimieren der Data-Warehouse-Speicherung
- Unterstützen hybrider transaktionaler analytischer Verarbeitung (HTAP) mit Azure Synapse Link
- Führen Sie End-to-End-Sicherheit mit Azure Synapse Analytics durch
- Durchführen von Stream Processing in Echtzeit mit Stream Analytics
- Erstellen einer Stream-Processing-Lösung mit Event Hubs und Azure Databricks
- Erstellen von Berichten mithilfe der Power-BI-Integration mit Azure Synapse Analytics
- Durchführen integrierter Machine-Learning-Prozesse in Azure Synapse Analytics

## Zielpublikum

Die primäre Zielgruppe für diesen Kurs sind Datenexperten, Datenarchitekten und Business-Intelligence-Experten, die etwas über Data Engineering und den Aufbau analytischer Lösungen mit Datenplattform-Technologien auf Microsoft Azure lernen möchten.

Die sekundäre Zielgruppe für diesen Kurs sind Datenanalysten und Datenwissenschaftler, die mit analytischen Lösungen arbeiten, die auf Microsoft Azure aufgebaut sind.

## Anforderungen

Erfolgreiche Teilnehmer beginnen diesen Kurs mit Kenntnissen über Cloud Computing und Kerndatenkonzepte sowie Berufserfahrung mit Datenlösungen.

Empfohlen wird das in folgenden Kursen erlangte Grundwissen:

- [Microsoft Azure Fundamentals \(Hands-on\) – Intensive Training \(«A900IC»\)](#)
- [Microsoft Azure Data Fundamentals – Intensive Training \(«DP900»\)](#)
- [Microsoft Azure Fundamentals – Flexible Training \(«AZ900V»\)](#)

## Zertifizierung

Dieses Intensive Training bereitet Sie vor auf:

- Prüfung: «DP-203: Data Engineering on Microsoft Azure» für die
- Zertifizierung: «Microsoft Certified: Azure Data Engineer Associate»

## Haben Sie Fragen oder möchten Sie einen Firmenkurs buchen?

Wir beraten Sie gerne unter 044 447 21 21 oder [info@digicomp.ch](mailto:info@digicomp.ch). Detaillierte Infos zu den Terminen finden Sie unter [www.digicomp.ch/weiterbildung-microsoft-technology/microsoft-azure/microsoft-certified-azure-data-engineer-associate/kurs-data-engineering-on-microsoft-azure-intensive-training-dp-203](https://www.digicomp.ch/weiterbildung-microsoft-technology/microsoft-azure/microsoft-certified-azure-data-engineer-associate/kurs-data-engineering-on-microsoft-azure-intensive-training-dp-203)