Data Engineering on Microsoft Azure. (DP-203)

Data Engineering on Microsoft Azure. (DP-203)

Seminar | 8 Termine verfügbar | Teilnahmebescheinigung

Präsenz / Virtual Classroom | 32 Unterrichtseinheiten | Online durchführbar

Seminarnummer: 29528 | Herstellernummer: MOC-DP-203

Stand: 19.03.2024. Alle aktuellen Informationen finden Sie unter https://akademie.tuv.com/s/29528

In diesem Seminar lernen Sie die Data Engineering-Muster und -Praktiken kennen, die sich auf die Arbeit mit Batch- und Echtzeit-Analyselösungen unter Verwendung von Azure-Datenplattformtechniken beziehen. Sie beginnen dabei mit Grundlagen der wichtigsten Computer- und Speichertechnologien, die zum Erstellen einer analytischen Lösung verwendet werden und lernen darauf aufbauende Lösungen kennen.

Nutzen

- Informationen über Rechen- und Speicheroptionen für Data Engineering-Workloads in Azure
- Entwerfen und Implementieren der Ausgabeebene
- Überlegungen zu Data Engineering verstehen
- Ausführen interaktiver Abfragen mit serverlosen SQL-Pools
- Durchsuchen, Transformieren und Laden von Daten in das Daten-Warehouse mit Apache Spark
- Datenexploration und -transformation in Azure Databricks
- Daten aufnehmen und in das Daten-Warehouse laden
- Transformieren von Daten mit Azure Data Factory- oder Azure Synapse-Pipelines
- Integrieren von Notebook-Daten in Azure Data Factory- oder Azure Synapse-Pipelines
- Optimieren der Abfrageleistung mit dedizierten SQL-Pools in Azure Synapse
- Analysieren und Optimieren der Daten-Warehouse-Speicher
- Unterstützung der Hybrid-Transaktionsanalysebearbeitung (HTAP) mit Azure Synapse Link
- End-to-End-Sicherheit mit Azure Synapse Analytics
- Echtzeit-Stream-Verarbeitung mit Stream Analytics
- Stream-Verarbeitungslösung mit Event Hubs und Azure Databricks
- Berichte mit Hilfe der Power BI-Integration mit Azure Synapse Analytics

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

- Integrierte maschinelle Lernprozesse in Azure Synapse Analytics

Zielgruppe

Die Zielgruppe für dieses Seminar sind Datenfachleute, Datenarchitekt:innen und Expert:innen für Geschäftsintelligenz, die mehr über Daten-Engineering und das Erstellen von Analyselösungen mit Hilfe von Datenplattformtechnologien in Microsoft Azure erfahren möchten. Eine weitere Zielgruppe für diesen Kurs sind Datenanalyst:innen und Datenwissenschaftler:innen, die mit auf Microsoft Azure basierenden Analyselösungen arbeiten.

Voraussetzungen

Teilnehmer:innen beginnen diesen Kurs mit vorhandenen Kenntnissen in Cloud Computing zum Beispiel aus dem Seminar AZ-900 (Azure Fundamentals) und Kerndatenkonzepten sowie Berufserfahrung mit Datenlösungen.

Inhalte des Seminars

Modul 1: Untersuchen von Rechen- und Speicheroptionen für Daten-Engineering-Workloads

Dieses Modul bietet einen Überblick über die Optionen der Azure-Computer- und Speichertechnologie, die Dateningenieur:innen zur Verfügung stehen, die analytische Workloads erstellen. In diesem Modul erfahren Sie, wie man das Datenmeer strukturiert und die Dateien für Explorations-, Streaming- und Batch-Workloads optimieren kann. Die Teilnehmer lernen, wie man das Datenmeer in Ebenen der Datenverfeinerung organisiert, während Dateien durch Stapel- und Stream-Verarbeitung transformiert werden. Anschließend erfahren Sie, wie man Indizes für Ihre Datensets wie CSV-, JSON- und Parkettdateien erstellen und diese für eine mögliche Beschleunigung von Abfragen und Workloads verwenden kann.

- Einführung in Azure Synapse Analytics
- Beschreiben von Azure Databricks
- Einführung in den Azure Datenmeer-Speicher
- Die Architektur des Datenmeeres beschreiben
- Arbeiten mit Datenströmen, mithilfe von Azure Stream Analytics

Modul 2: Entwerfen und Implementieren der Ausgabeebenen

In diesem Modul erfahren Sie, wie Datenspeicher in einem modernen Daten-Warehouse entworfen und implementiert werden, um analytische Workloads zu optimieren. Sie lernen, wie ein mehrdimensionales Schema zum Speichern von Fakten- und Dimensionsdaten entworfen wird. Anschließend lernen Sie, wie langsam ändernde Dimensionen durch inkrementelles Laden von Daten aus Azure Data Factory gefüllt

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

werden.

- Entwerfen eines mehrdimensionales Schemas, um analytische Workloads zu optimieren
- Codefreie Transformation im Maßstab mit Azure Data Factory
- Sich langsam ändernde Dimensionen in Azure Synapse Analytics-Pipelines füllen

Modul 3: Überlegungen zum Daten-Engineering für Quelldateien

In diesem Modul werden Überlegungen zum Daten-Engineering erläutert, die beim Laden von Daten in ein modernes Daten-Warehouse häufig vorkommen, welches analytisch geladen wird aus in einem Azure Datenmeer von gespeicherten Dateien. In diesem Zusammenhang werden Ihnen auch die Sicherheitsaspekte beim Speichern von im Datenmeer gespeicherten Dateien vermittelt.

- Entwerfen Sie ein modernes Daten-Warehouse mit Azure Synapse Analytics
- Sichern Sie ein Daten-Warehouse in Azure Synapse Analytics

Modul 4: Interaktive Abfragen mit serverlosen SQL-Pools von Azure Synapse Analytics

In diesem Modul werden Überlegungen zum Daten-Engineering erläutert, die beim Laden von Daten in ein modernes Daten-Warehouse häufig vorkommen, welches analytisch geladen wird aus in einem Azure Datenmeer von gespeicherten Dateien. In diesem Zusammenhang werden Ihnen auch die Sicherheitsaspekte beim Speichern von im Datenmeer gespeicherten Dateien vermittelt.

- Erkunden Sie die Funktionen der serverlosen SQL-Poolsvon Azure Synapse
- Abfragen von Daten im Meer mit Hilfe von serverlosen SQL-Pools von Azure Synapse
- Erstellen Sie Metadatenobjekte in serverlosen SQL-Pools von Azure Synapse
- Sichern Sie Daten und verwalten Sie Benutzer in serverlosen SQL-Pools von Azure Synapse

Modul 5: Durchsuchen, Transformieren und Laden von Daten in das Daten-Warehouse mit Apache Spark

In diesem Modul erfahren Sie, wie man in einem Datenmeer gespeicherte Daten untersucht, die Daten transformiert und Daten in einen relationalen Datenspeicher lädt. Sie werden in dem Seminar Parkett- und JSON-Dateien untersuchen und Techniken verwenden, um JSON-Dateien mit hierarchischen Strukturen abzufragen und zu transformieren. Anschließend verwenden Sie Apache Spark, um Daten in das Daten-Warehouse zu laden und Parkettdaten im Datenmeer zu verknüpfen mit Daten im dedizierten SQL-Pool.

- Grundlegendes zum Engineering großer Datenmengen mit Apache Spark in Azure Synapse Analytics
- Daten mit Apache Spark-Notebooks in Azure Synapse Analytics aufnehmen
- Daten mit Data Frames in Apache Spark-Pools in Azure Synapse Analytics transformieren
- SQL- und Apache Spark-Pools in Azure Synapse Analytics integrieren

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

Modul 6: Datenexploration und -transformation in Azure Databricks

In diesem Modul erfahren Sie, wie man mit verschiedenen Apache Spark Data Frame-Methoden Daten in Azure Databricks untersuchen und transformieren kann. Teilnehmer:innen lernen, wie man Standard-Data Frame-Methoden zum Erkunden und Transformieren von Daten ausführen kann. Sie lernen auch, wie man erweiterte Aufgaben ausführen, z. B. doppelte Daten entfernen, Datums- / Zeitwerte bearbeiten, Spalten umbenennen und Daten aggregieren kann.

- Beschreiben von Azure Databricks
- Lesen und Schreiben von Daten in Azure Databricks
- Arbeiten mit Data Frames in Azure Databricks
- Arbeiten mit erweiterten Data Frames-Methoden in Azure Databricks

Modul 7: Daten aufnehmen und in das Daten- Warehouse laden

In diesem Modul lernen Sie, wie man Daten über T-SQL-Skripte und Synapse Analytics-Integrationspipelines in das Daten-Warehouse aufnimmt. Sie erfahren, wie Daten mit PolyBase und COPY unter Anwendung von TSQL in dedizierte Synapse-SQL-Pools geladen werden. Außerdem zeigen wir Ihnen, wie das Workload-Management, zusammen mit einer Kopieraktivität in einer Azure Synapse-Pipeline für die Datenaufnahme im Petabyte-Bereich verwendet wird.

- Verwenden Sie bewährte Praktiken zum Laden von Daten in Azure Synapse Analytics
- Aufnahme im Petabyte-Bereich mit Azure Data Factory

Modul 8: Transformieren von Daten mit Azure Data Factory- oder Azure Synapse-Pipelines

In diesem Modul lernen Teilnehmer:innen, wie man Datenintegrations-Pipelines erstellt, um sie aus mehreren Datenquellen aufzunehmen, Daten mit Hilfe von Mapping-Datenflüssen zu transformieren und Daten in eine oder mehrere Datensenken zu verschieben.

- Datenintegration mit Azure Data Factory- oder Azure Synapse-Pipelines
- Codefreie Transformation im Maßstab mit Azure Data Factory- oder Azure Synapse-Pipelines

Modul 9: Koordination von Verschieben und Transformieren von Daten in Azure Synapse-Pipelines

In diesem Modul erfahren Sie, wie man verknüpfte Services erstellt und die Datenverschiebung und -transformation mit Hilfe von Notebooks in Azure Synapse-Pipelines koordiniert.

- Orchestrieren der Datenverschiebung und -transformation in Azure Data Factory

Modul 10: Optimieren Sie die Abfrageleistung mit dedizierten SQL-Pools in Azure Synapse

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

Im zehnten Modul lernen Sie Strategien zur Optimierung der Datenspeicherung und -verarbeitung bei Verwendung dedizierter SQL-Pools in Azure Synapse Analytics. Sie können die Abfrageleistung optimieren und verbessern, durch Entwicklerfunktionen wie Fenster- und HyperLogLog-Funktionen und bewährte Methoden zum Laden von Daten.

- Optimierung der Leistung der Abfragen des Daten-Warehouses in Azure Synapse Analytics
- Kenntnis der Data Warehouse Entwickler-Features von Azure Synapse Analytics

Modul 11: Analyse und Optimierung der Daten Warehouse Speicherung

In Modul Nummer 11 lernen Sie, wie die Datenspeicherung von Azure Synapse den dezidierten SQL analysiert und optimiert. Sie werden Techniken erlernen, um die Verwendung des Tabellenbereichs und die Speicherdetails des Spaltenspeichers zu verstehen. Außerdem werden Ihnen die Speicheranforderungen zwischen identischen Tabellen, die unterschiedliche Datentypen verwenden, erklärt. Sie können dann beobachten, welche Auswirkungen materialisierte Ansichten haben, wenn sie anstelle komplexer Abfragen ausgeführt werden, und lernen, wie durch Optimierung der Löschvorgänge eine umfangreiche Protokollierung vermieden wird.

- Analysieren und Optimieren der Daten-Warehouse-Speicher in Azure Synapse Analytic

Modul 12: Unterstützung von HTAP (Hybrid Transactional Analytical Processing) mit Azure Synapse Links

In diesem Modul erfahren Teilnehmer:innen, wie Azure Synapse Link die nahtlose Verbindung eines Azure Cosmos DB-Kontos mit einem Synapse-Arbeitsbereich ermöglicht. Sie lernen zu verstehen, wie man die Synapse-Verknüpfung aktiviert und konfiguriert und wie der Azure Cosmos DB-Analysespeicher mit Apache Spark und serverlosem SQL abgefragt wird.

- Entwerfen hybrider Transaktions- und Analyseverarbeitung mit Azure Synapse Analytics
- Konfiguration von Azure Synapse Link mit Azure Cosmos DB
- Abfrage von Azure Cosmos DB mit Apache Spark-Pools ab
- Abfrage von Azure Cosmos DB mit serverlosen SQL-Pools ab

Modul 13: End-to-End-Sicherheit mit Azure Synapse Analytics

Unser Seminar lehrt Ihnen, wie ein Synapse Analytics-Arbeitsbereich und seine unterstützende Infrastruktur gesichert werden. Sie wissen den SQL Active Directory-Administrator zu beobachten, verwalten IP-Firewall-Regeln und Geheimnisse mit Azure Key Vault und greifen auf diese Geheimnisse über einen mit Key Vault verknüpften Service und Pipeline-Aktivitäten zu. Sie werden verstehen, wie Sicherheit auf Spalten- und Zeilenebene und dynamische Datenmaskierung implementiert werden, bei Verwendung von dedizierten SQL-Pools.

- Sicherung vom Daten-Warehouse in Azure Synapse Analytics

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

- Konfigurieren und Verwalten von Geheimnissen in Azure Key Vault
- Implementieren von Compliance-Kontrollen für sensible Daten

Modul 14: Echtzeit-Stream-Verarbeitung mit Stream Analytics

Sie lernen, wie Streaming-Daten mit Azure Stream Analytics verarbeitet werden. Wir nehmen Fahrzeugtelemetrie-Daten in Event Hubs auf und verarbeiten sie dann in Echtzeit mit Hilfe verschiedener Fensterfunktionen in Azure Stream Analytics. Die Daten werden dann an Azure Synapse Analytics ausgegeben. Um den Durchsatz zu erhöhen, zeigen wir Ihnen den Stream Analytics-Job optimal zu skalieren.

Zuverlässiges Messaging für Big Data-Anwendungen gelingt Ihnen mit Hilfe von Azure Event Hubs. Darüber hinaus ermöglicht es Azure Stream Analytics Datenströme aufzunehmen.

Modul 15: Erstellen einer Stream-Verarbeitungslösung mit Event Hubs und Azure Databricks

Unsere Expert:innen zeigen Ihnen, wie Streaming-Daten mit Event Hubs und Spark Structured Streaming in Azure Databricks maßstabsgetreu erfasset und verarbeitet werden können. Dabei lernen Sie die wichtigsten Funktionen und Verwendungszwecke von Structured Streaming kennen. Gemeinsam implementieren wir Schiebefenster, um Datenblöcke zu aggregieren und wenden Wasserzeichen an, um veraltete Daten zu entfernen. Schließlich stellen wir eine Verbindung zu Event Hubs her, um Streams zu lesen und zu schreiben.

- Verarbeiten von Streaming-Daten mit strukturiertem Azure Databricks-Streaming

Modul 16: Erstellen von Berichten mit Hilfe der Power BI-Integration mit Azure Synapse Analytics

In Modul 16 lernen Sie, wie sich Power BI in Ihren Synapse-Arbeitsbereich integrieren lässt, um Berichte in Power BI zu erstellen. Sie erstellen eine neue Datenquelle und einen neuen Power BI-Bericht in Synapse Studio. Anschließend lernen Sie, wie man die Abfrageleistung durch materialisierte Ansichten und Zwischenspeichern von Ergebnismengen verbessern kann. Darauf aufbauend werden wir das Datenmeer mit serverlosen SQL-Pools erkunden und Visualisierungen für diese Daten in Power BI erstellen.

- Power BI Reports mit Hilfe der Integration in Azure Synapse Analytics

Modul 17: Integrierte maschinelle Lernprozesse in Azure Synapse Analytics ausführen

In diesem Modul wird die integrierte, durchgängige Erfahrung mit Azure maschinellen Lernen und Azure kognitiven Services in Azure Synapse Analytics erläutert. Sie erfahren, wie Sie einen Azure Synapse Analytics-Arbeitsbereich mithilfe eines verknüpften Services mit einem Azure Machine Learning-Arbeitsbereich verbinden und anschließend ein automatisiertes ML-Experiment auslösen, das Daten aus einer Spark-Tabelle verwendet. Außerdem erfahren Sie, wie man trainierte Modelle aus Azure-maschiniertem Lernen oder Azure kognitiven Services verwenden, um Daten in einer SQL-Pooltabelle anzureichern und anschließend mit Hilfe von Power BI Prognoseergebnisse bereitzustellen.

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

- Verwenden von integrierten maschinellen Lernprozessn in Azure Synapse Analytics

Wichtige Hinweise

ACHTUNG: Der Seminartermin ab 27.7 wird als gekürztes Training mit komprimierter Inhaltsvermittlung in 3 Tagen durchgeführt!

Terminübersicht und Buchung

Buchen Sie Ihren Wunschtermin jetzt direkt online unter https://akademie.tuv.com/s/29528 und profitieren Sie von diesen Vorteilen:

- Schneller Buchungsvorgang
- Persönliches Kundenkonto
- Gleichzeitige Buchung für mehrere Teilnehmer:innen

Alternativ können Sie das Bestellformular verwenden, um via Fax oder E-Mail zu bestellen.

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.