

**EDUCACIÓN **

**Data Engineering on  
Microsoft Azure (DP-203T00)**

Programa de Estudio

# Data Engineering on Microsoft Azure (DP-203T00)

Aprende sobre la ingeniería de datos en base al trabajo con soluciones analíticas en tiempo real utilizando tecnologías de plataforma de datos de Azure.

## Contenido del curso

---

Los estudiantes comenzarán por comprender las tecnologías básicas de computación y almacenamiento que se utilizan para construir una solución analítica. Los estudiantes aprenderán a explorar de forma interactiva los datos almacenados en archivos en un lago de datos. Aprenderán las diversas técnicas de ingestión que se pueden usar para cargar datos utilizando la capacidad de Apache Spark que se encuentra en Azure Synapse Analytics o Azure Databricks, o cómo ingerir usando Azure Data Factory o las canalizaciones de Azure Synapse. También aprenderán las diversas formas en que pueden transformar los datos utilizando las mismas tecnologías que se utilizan para ingerir datos. Comprenderán la importancia de implementar la seguridad para garantizar que los datos estén protegidos en reposo o en tránsito. Luego, el estudiante mostrará cómo crear un sistema analítico en tiempo real para crear soluciones analíticas en tiempo real

## Requisitos

---

Conocimientos básicos de Azure o haber realizado el curso Microsoft Azure Fundamentals AZ-900

Conocimientos básicos de datos de Microsoft Azure o haber realizado el curso Microsoft Azure Data Fundamentals DP-900

## Modalidad de cursado

---

Puedes tomar este curso en modalidad presencial o modalidad online - en vivo

# ¿Qué aprenderás?

---

- Explorar las opciones de procesamiento y almacenamiento para cargas de trabajo de ingeniería de datos en Azure
- Ejecutar consultas interactivas utilizando grupos SQL sin servidor
- Realizar exploración y transformación de datos en Azure Databricks
- Explorar, transforme y cargue datos en el almacén de datos con Apache Spark
- Aprender sobre ingesta y carga de datos en el almacén de datos
- Transformar datos con Azure Data Factory o Azure Synapse Pipelines
- Integrar datos de portátiles con Azure Data Factory o Azure Synapse Pipelines
- Admitir el procesamiento analítico transaccional híbrido (HTAP) con Azure Synapse Link
- Realizar la seguridad de un extremo a otro con Azure Synapse Analytics
- Realizar el procesamiento de transmisiones en tiempo real con Stream Analytics
- Crear una solución de procesamiento de transmisión con Event Hubs y Azure Databricks



# Plan de Estudios



## 1. Explore las opciones de computación y almacenamiento para cargas de trabajo de ingeniería de datos

---

Introducción a Azure Synapse Analytics

Describir Azure Databricks

Introducción al almacenamiento de Azure Data Lake

Describir la arquitectura de Delta Lake

Trabajar con flujos de datos mediante Azure Stream Analytics

Explore las opciones de computación y almacenamiento para cargas de trabajo de ingeniería de datos

Combine la transmisión y el procesamiento por lotes con una única canalización

Organice el lago de datos en niveles de transformación de archivos

Indexación de almacenamiento de lago de datos para la aceleración de consultas y cargas de trabajo

Describir Azure Synapse Analytics

Describir Azure Databricks

Describir el almacenamiento de Azure Data Lake

Describir la arquitectura de Delta Lake

Describir Azure Stream Analytics

## 2. Ejecute consultas interactivas con grupos de SQL sin servidor de Azure Synapse Analytics

---

Explore las capacidades de los grupos SQL sin servidor de Azure Synapse

Consultar datos en el lago mediante grupos de SQL sin servidor de Azure Synapse

Crear objetos de metadatos en grupos SQL sin servidor de Azure Synapse

Proteja los datos y administre a los usuarios en los grupos SQL sin servidor de Azure Synapse

Ejecute consultas interactivas utilizando grupos de SQL sin servidor

Consultar datos de Parquet con grupos SQL sin servidor

Crea tablas externas para archivos Parquet y CSV

Cree vistas con grupos de SQL sin servidor

Acceso seguro a los datos en un lago de datos cuando se utilizan grupos de SQL sin servidor

Configure la seguridad del lago de datos mediante el control de acceso basado en roles (RBAC) y la lista de control de acceso

Comprender las capacidades de los grupos de SQL sin servidor de Azure Synapse

Consultar datos en el lago mediante grupos de SQL sin servidor de Azure Synapse

Crear objetos de metadatos en grupos SQL sin servidor de Azure Synapse

Proteja los datos y administre a los usuarios en los grupos SQL sin servidor de Azure Synapse

### 3. Exploración y transformación de datos en Azure Databricks

---

Describir Azure Databricks

Leer y escribir datos en Azure Databricks

Trabajar con DataFrames en Azure Databricks

Trabajar con métodos avanzados de DataFrames en Azure Databricks

Exploración y transformación de datos en Azure Databricks

Use DataFrames en Azure Databricks para explorar y filtrar datos

Almacene en caché un DataFrame para consultas posteriores más rápidas

Eliminar datos duplicados

Manipular valores de fecha / hora

Eliminar y cambiar el nombre de las columnas DataFrame

Agregar datos almacenados en un DataFrame

Describir Azure Databricks

Leer y escribir datos en Azure Databricks

Trabajar con DataFrames en Azure Databricks

Trabajar con métodos avanzados de DataFrames en Azure Databricks

### 4. Explore, transforme y cargue datos en el almacén de datos usando Apache Spark

---

Comprender la ingeniería de big data con Apache Spark en Azure Synapse Analytics

Ingesta datos con los cuadernos de Apache Spark en Azure Synapse Analytics

Transforme datos con DataFrames en Apache Spark Pools en Azure Synapse Analytics

Integrar grupos de SQL y Apache Spark en Azure Synapse Analytics

Explore, transforme y cargue datos en el almacén de datos usando Apache Spark

Realizar exploración de datos en Synapse Studio

Ingesta datos con cuadernos Spark en Azure Synapse Analytics

Transforme datos con DataFrames en grupos de Spark en Azure Synapse Analytics

Integrar grupos de SQL y Spark en Azure Synapse Analytics

Describir la ingeniería de big data con Apache Spark en Azure Synapse Analytics

Ingesta datos con los cuadernos de Apache Spark en Azure Synapse Analytics

Transforme datos con DataFrames en Apache Spark Pools en Azure Synapse Analytics

Integrar grupos de SQL y Apache Spark en Azure Synapse Analytics

## 5. Ingesta y carga datos en el almacén de datos

---

Utilice las mejores prácticas de carga de datos en Azure Synapse Analytics

Ingestión a escala de petabytes con Azure Data Factory

Ingesta y carga datos en el almacén de datos

Realice la ingestión a escala de petabytes con Azure Synapse Pipelines

Importe datos con PolyBase y COPY usando T-SQL

Utilice las mejores prácticas de carga de datos en Azure Synapse Analytics

Después de completar este módulo, los estudiantes podrán:

Utilice las mejores prácticas de carga de datos en Azure Synapse Analytics

Ingestión a escala de petabytes con Azure Data Factory

## 6. Transformar datos con Azure Data Factory o Azure Synapse Pipelines

---

Integración de datos con Azure Data Factory o Azure Synapse Pipelines

Transformación sin código a escala con Azure Data Factory o Azure Synapse Pipelines

Transformar datos con Azure Data Factory o Azure Synapse Pipelines

Ejecute transformaciones sin código a escala con Azure Synapse Pipelines

Cree una canalización de datos para importar archivos CSV con formato deficiente

Crear flujos de datos de mapeo

Después de completar este módulo, los estudiantes podrán:

Realice la integración de datos con Azure Data Factory

Realice una transformación sin código a escala con Azure Data Factory

## 7. Orquestar el movimiento y la transformación de datos en Azure Synapse Pipelines

---

Organice el movimiento y la transformación de datos en Azure Data Factory

Orquestar el movimiento y la transformación de datos en Azure Synapse Pipelines

Integre datos de portátiles con Azure Data Factory o Azure Synapse Pipelines

Después de completar este módulo, los estudiantes podrán:

Organice el movimiento y la transformación de datos en Azure Synapse Pipelines



## 8. Seguridad de un extremo a otro con Azure Synapse Analytics

---

Proteger un almacén de datos en Azure Synapse Analytics

Configurar y administrar secretos en Azure Key Vault

Implementar controles de cumplimiento para datos confidenciales

Seguridad de un extremo a otro con Azure Synapse Analytics

Asegure la infraestructura de soporte de Azure Synapse Analytics

Asegure el área de trabajo de Azure Synapse Analytics y los servicios administrados

Proteger los datos del área de trabajo de Azure Synapse Analytics

Después de completar este módulo, los estudiantes podrán:

Proteger un almacén de datos en Azure Synapse Analytics

Configurar y administrar secretos en Azure Key Vault

Implementar controles de cumplimiento para datos confidenciales

## 9. Soporte del procesamiento analítico transaccional híbrido (HTAP) con Azure Synapse Link

---

Diseño procesamiento transaccional y analítico híbrido con Azure Synapse Analytics

Configurar Azure Synapse Link con Azure Cosmos DB

Consultar Azure Cosmos DB con grupos de Apache Spark

Consultar Azure Cosmos DB con grupos de SQL sin servidor

Soporte de procesamiento analítico transaccional híbrido (HTAP) con Azure Synapse Link

Configurar Azure Synapse Link con Azure Cosmos DB

Consultar Azure Cosmos DB con Apache Spark para Synapse Analytics

Consultar Azure Cosmos DB con un grupo de SQL sin servidor para Azure Synapse Analytics

Diseño procesamiento transaccional y analítico híbrido con Azure Synapse Analytics

Configurar Azure Synapse Link con Azure Cosmos DB

Consultar Azure Cosmos DB con Apache Spark para Azure Synapse Analytics

Consultar Azure Cosmos DB con SQL sin servidor para Azure Synapse Analytics

## 10. Procesamiento de transmisión en tiempo real con Stream Analytics

---

Habilite la mensajería confiable para aplicaciones de Big Data con Azure Event Hubs



Trabajar con flujos de datos mediante Azure Stream Analytics

Ingesta flujos de datos con Azure Stream Analytics

Procesamiento de transmisión en tiempo real con análisis de transmisión

Utilice Stream Analytics para procesar datos en tiempo real de Event Hubs

Utilice las funciones de ventana de Stream Analytics para crear agregados y resultados en Synapse Analytics

Escale el trabajo de Azure Stream Analytics para aumentar el rendimiento mediante la partición

Repartir la entrada de flujo para optimizar la paralelización

Habilite la mensajería confiable para aplicaciones de Big Data con Azure Event Hubs

Trabajar con flujos de datos mediante Azure Stream Analytics

Ingesta flujos de datos con Azure Stream Analytics

## **11. Cree una solución de procesamiento de transmisión con Event Hubs y Azure Databricks**

---

Procesar datos de transmisión con transmisión estructurada de Azure Databricks

Cree una solución de procesamiento de transmisión con Event Hubs y Azure Databricks

Explore las características y usos clave de la transmisión estructurada

Transmita datos desde un archivo y escríbalos en un sistema de archivos distribuido

Utilice ventanas deslizantes para agregar fragmentos de datos en lugar de todos los datos

Aplicar marcas de agua para eliminar datos obsoletos

Conectarse a transmisiones de lectura y escritura de Event Hubs

Procesar datos de transmisión con transmisión estructurada de Azure Databricks

# EDUCACIÓN IT

*Centro de Capacitación y Desarrollo Profesional*



Lavalle 648 Piso 8, Microcentro, CABA

TEL\_PRINCIPAL

[info@educacionit.com](mailto:info@educacionit.com)

EducaciónIT. Copyright 2005-2021