

Data Engineering on Microsoft Azure

Temario

1. Módulo 1: Explorar las opciones de computación y almacenamiento para cargas de trabajo de ingeniería de datos
 - 1.1 Introducción a Azure Synapse Analytics
 - 1.2 Describir Azure Databricks
 - 1.3 Introducción al almacenamiento de Azure Data Lake
 - 1.4 Describir la arquitectura de Delta Lake
 - 1.5 Trabajar con flujos de datos mediante Azure Stream Analytics
Laboratorio: Explorar las opciones de computación y almacenamiento para cargas de trabajo de ingeniería de datos
 - Combinar la transmisión y el procesamiento por lotes con una única pipeline
 - Organizar el data lake en niveles de transformación de archivos
 - Indexación de almacenamiento de data lake para la aceleración de consultas y cargas de trabajo

2. Módulo 2: Diseño e implementación de la capa de servicio
 - 2.1 Diseñar un esquema multidimensional para optimizar las cargas de trabajo analíticas
 - 2.2 Transformación sin código a escala con Azure Data Factory
 - 2.3 Rellenar dimensiones que cambian lentamente en las pipelines de Azure Synapse Analytics
Laboratorio: Diseño e implementación de la capa de servicio
 - Diseñar un esquema en estrella para cargas de trabajo analíticas
 - Rellenar dimensiones que cambian lentamente con Azure Data Factory y mapeo de flujos de datos

3. Módulo 3: Consideraciones de ingeniería de datos para archivos fuente
 - 3.1 Diseñar un almacén de datos moderno con Azure Synapse Analytics
 - 3.2 Proteger un almacén de datos en Azure Synapse Analytics
Laboratorio: Consideraciones de ingeniería de datos
 - Administrar archivos en un data lake de Azure
 - Protección de archivos almacenados en un data lake de Azure

4. Módulo 4: Ejecutar consultas interactivas con grupos de SQL sin servidor de Azure Synapse Analytics
 - 4.1 Explorar las capacidades de los grupos SQL sin servidor de Azure Synapse
 - 4.2 Consultar datos en el lake mediante grupos de SQL sin servidor de Azure Synapse
 - 4.3 Crear objetos de metadatos en grupos SQL sin servidor de Azure Synapse
 - 4.4 Proteger los datos y administrar a los usuarios en los grupos SQL sin servidor de Azure Synapse
Laboratorio: Ejecutar consultas interactivas utilizando grupos de SQL sin servidor
 - Consultar datos de Parquet con grupos SQL sin servidor
 - Crear tablas externas para archivos Parquet y CSV
 - Crear vistas con grupos de SQL sin servidor

- Acceso seguro a los datos en un data lake cuando se utilizan grupos de SQL sin servidor
 - Configurar la seguridad del data lake mediante el control de acceso basado en roles (RBAC) y la lista de control de acceso
5. Módulo 5: Explorar, transformar y cargar datos en el almacén de datos usando Apache Spark
- 5.1 Comprender la ingeniería de big data con Apache Spark en Azure Synapse Analytics
 - 5.2 Ingestar datos con los cuadernos de Apache Spark en Azure Synapse Analytics
 - 5.3 Transformar datos con DataFrames en Apache Spark Pools en Azure Synapse Analytics
 - 5.4 Integrar grupos de SQL y Apache Spark en Azure Synapse Analytics
- Laboratorio: Explorar, transformar y cargar datos en el almacén de datos usando Apache Spark
- Realizar exploración de datos en Synapse Studio
 - Ingestar datos con cuadernos Spark en Azure Synapse Analytics
 - Transformar datos con DataFrames en grupos de Spark en Azure Synapse Analytics
 - Integrar grupos de SQL y Spark en Azure Synapse Analytics
6. Módulo 6: Exploración y transformación de datos en Azure Databricks
- 6.1 Describir Azure Databricks
 - 6.2 Leer y escribir datos en Azure Databricks
 - 6.3 Trabajar con DataFrames en Azure Databricks
 - 6.4 Trabajar con métodos avanzados de DataFrames en Azure Databricks
- Laboratorio: Exploración y transformación de datos en Azure Databricks
- Usar DataFrames en Azure Databricks para explorar y filtrar datos
 - Almacenar 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
7. Módulo 7: Ingesta y carga de datos en el almacén de datos.
- 7.1 Utilizar las mejores prácticas de carga de datos en Azure Synapse Analytics
 - 7.2 Ingestión a escala de petabytes con Azure Data Factory
- Laboratorio: Ingesta y carga de datos en el almacén de datos
- Realizar la ingestión a escala de petabytes con Azure Synapse Pipelines
 - Importar datos con PolyBase y COPY usando T-SQL
 - Utilizar las mejores prácticas de carga de datos en Azure Synapse Analytics
8. Módulo 8: Transformar datos con Azure Data Factory o Azure Synapse Pipelines
- 8.1 Integración de datos con Azure Data Factory o Azure Synapse Pipelines
 - 8.2 Transformación sin código a escala con Azure Data Factory o Azure Synapse Pipelines
- Laboratorio: Transformar datos con Azure Data Factory o Azure Synapse Pipelines
- Ejecutar transformaciones sin código a escala con Azure Synapse Pipelines

- Crear un pipeline de datos para importar archivos CSV con formato deficiente
 - Crear flujos de datos de mapeo
9. Módulo 9: Orquestar el movimiento y la transformación de datos en Azure Synapse Pipelines
- 9.1 Organizar el movimiento y la transformación de datos en Azure Data Factory
- Laboratorio: Orquestar el movimiento y la transformación de datos en Azure Synapse Pipelines
- Integrar datos de portátiles con Azure Data Factory o Azure Synapse Pipelines
10. Módulo 10: Optimizar el rendimiento de las consultas con grupos de SQL dedicados en Azure Synapse
- 10.1 Optimizar el rendimiento de las consultas del almacén de datos en Azure Synapse Analytics
- 10.2 Comprender las características para desarrolladores de almacenamiento de datos de Azure Synapse Analytics
- Laboratorio: Optimizar el rendimiento de las consultas con grupos de SQL dedicados en Azure Synapse
- Comprender las características para desarrolladores de Azure Synapse Analytics
 - Optimizar el rendimiento de las consultas del almacén de datos en Azure Synapse Analytics
 - Mejorar el rendimiento de las consultas
11. Módulo 11: Analizar y optimizar el almacenamiento del data warehouse
- 11.1 Analizar y optimizar el almacenamiento del data warehouse de datos en Azure Synapse Analytics
- Laboratorio: Analizar y optimizar el almacenamiento del data warehouse
- Comprobar si hay datos sesgados y uso de espacio
 - Comprender los detalles de almacenamiento de la tienda de columnas
 - Estudiar el impacto de las vistas materializadas
 - Explorar las reglas para operaciones mínimamente registradas
12. Módulo 12: Soporte del procesamiento analítico transaccional híbrido (HTAP) con Azure Synapse Link
- 12.1 Diseñar procesamiento transaccional y analítico híbrido con Azure Synapse Analytics
- 12.2 Configurar Azure Synapse Link con Azure Cosmos DB
- 12.3 Consultar Azure Cosmos DB con grupos de Apache Spark
- 12.4 Consultar Azure Cosmos DB con grupos de SQL sin servidor
- Laboratorio: 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
13. Módulo 13: Seguridad de un extremo a otro con Azure Synapse Analytics

- 13.1 Proteger un almacén de datos en Azure Synapse Analytics
 - 13.2 Configurar y administrar secretos en Azure Key Vault
 - 13.3 Implementar controles de cumplimiento para datos confidenciales
 - Laboratorio: Seguridad de un extremo a otro con Azure Synapse Analytics
 - Asegurar la infraestructura de soporte de Azure Synapse Analytics
 - Asegurar el área de trabajo de Azure Synapse Analytics y los servicios administrados
 - Proteger los datos del área de trabajo de Azure Synapse Analytics
14. Módulo 14: Procesamiento de transmisión en tiempo real con Stream Analytics
- 14.1 Habilitar la mensajería confiable para aplicaciones de Big Data con Azure Event Hubs
 - 14.2 Trabajar con flujos de datos mediante Azure Stream Analytics
 - 14.3 Ingesta flujos de datos con Azure Stream Analytics
 - Laboratorio: Procesamiento de transmisión en tiempo real con análisis de transmisión
 - Utilizar Stream Analytics para procesar datos en tiempo real de Event Hubs
 - Utilizar las funciones de ventana de Stream Analytics para crear agregados y resultados en Synapse Analytics
 - Escalar el trabajo de Azure Stream Analytics para aumentar el rendimiento mediante la partición
 - Repartir la entrada de flujo para optimizar la paralelización
15. Módulo 15: Crear una solución de procesamiento de transmisión con Event Hubs y Azure Databricks
- 15.1 Procesar datos de streaming con transmisión estructurada de Azure Databricks
 - Laboratorio: Crear una solución de procesamiento de transmisión con Event Hubs y Azure Databricks
 - Explorar las características y usos clave de la transmisión estructurada
 - Transmitir datos desde un archivo y escribirlos en un sistema de archivos distribuido
 - Utilizar 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
16. Módulo 16: Generar informes mediante la integración de Power BI con Azure Synapse Analytics
- 16.1 Crear informes con Power BI utilizando su integración con Azure Synapse Analytics
 - Laboratorio: Generar informes utilizando la integración de Power BI con Azure Synapse Analytics
 - Integrar un área de trabajo de Azure Synapse y Power BI
 - Optimizar la integración con Power BI
 - Mejorar el rendimiento de las consultas con vistas materializadas y almacenamiento en caché de conjuntos de resultados
 - Visualizar datos con SQL sin servidor y crear un informe de Power BI

17. Módulo 17: Realizar procesos integrados de aprendizaje automático en Azure Synapse Analytics

17.1 Utilizar el proceso de aprendizaje automático integrado en Azure Synapse Analytics

Laboratorio: Realizar procesos integrados de aprendizaje automático en Azure Synapse Analytics

- Crear un servicio vinculado de Azure Machine Learning
- Activar un experimento de Auto ML con datos de una tabla Spark
- Enriquecer los datos utilizando modelos entrenados
- Ofrecer resultados de predicción con Power BI

18. RFCs Adicionales en PHP 8.2