

# TEMARIO

## Data Engineering on Microsoft Azure



**SIV & DB CLOUD**

EXPERIENCIA Y TECNOLOGIA

### CONTÁCTENOS

 +57 316 3956090

 [contactenos@siv.com.co](mailto:contactenos@siv.com.co)

 [www.siv.com.co](http://www.siv.com.co)

 +57 315 2653920

 [comercial@siv.com.co](mailto:comercial@siv.com.co)

 [www.dbcloud.co](http://www.dbcloud.co)

## Data Engineering on Microsoft Azure

### Examen DP-203

#### Descripción del curso

En este curso, el alumno aprenderá sobre la ingeniería de datos en lo que se refiere al trabajo con soluciones analíticas por lotes y en tiempo real usando tecnologías de plataforma de datos de Azure. Los alumnos empezarán conociendo las tecnologías de proceso y almacenamiento básicas que se usan para crear una solución analítica. También aprenderán a explorar interactivamente los datos almacenados en archivos de un lago de datos. Conocerán las diversas técnicas de ingesta que se pueden usar para cargar datos mediante la funcionalidad de Apache Spark incluida en Azure Synapse Analytics o Azure Databricks, o cómo ingerir mediante canalizaciones de Azure Data Factory o Azure Synapse. Los alumnos también conocerán las distintas formas en que pueden transformar los datos mediante las mismas tecnologías que se usan para ingerir datos. Comprenderán la importancia de implementar seguridad para garantizar que los datos (en reposo o en tránsito) están protegidos. Tras ello, se les explicará cómo crear un sistema analítico en tiempo real para crear soluciones analíticas en tiempo real.

#### Perfil de la audiencia

La audiencia principal de este curso son profesionales de datos, arquitectos de datos y profesionales de business intelligence que quieren obtener información sobre la ingeniería de datos y la creación de soluciones analíticas mediante las tecnologías de plataforma de datos que hay en Microsoft Azure. La audiencia secundaria de este curso son analistas de datos y científicos de datos que trabajan con soluciones analíticas basadas en Microsoft Azure.

#### Prerrequisitos

Los alumnos aptos comienzan este curso con conocimientos de informática en la nube y fundamentos de datos, y experiencia profesional con soluciones de datos. Realizando en concreto:

- AZ-900: Fundamentos de Azure
- DP-900: Fundamentos de datos en Microsoft Azure

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## DESCRIPCION MODULOS DE CAPACITACION

### Módulo 1: Exploración de las opciones de proceso y almacenamiento de las cargas de trabajo de ingeniería de datos

En este módulo se proporciona información general sobre las opciones de tecnología de proceso y almacenamiento de Azure que hay disponibles para los ingenieros de datos que crean cargas de trabajo analíticas. En este módulo se enseña a estructurar el lago de datos y a optimizar los archivos para las cargas de trabajo de exploración, secuencias y lotes. El alumno aprenderá a organizar el lago de datos en niveles de refinamiento de datos a medida que transforman los archivos a través del procesamiento por lotes y de secuencias. Luego, aprenderán a crear índices en sus conjuntos de datos, como archivos CSV, JSON y Parquet, y a usarlos para una posible aceleración de las consultas y las cargas de trabajo.

#### Lecciones

- Introducción a Azure Synapse Analytics
- Descripción de Azure Databricks
- Introducción a Azure Data Lake Storage
- Descripción de la arquitectura de Delta Lake
- Trabajo con secuencias de datos mediante Azure Stream Analytics

#### Laboratorio: Exploración de las opciones de proceso y almacenamiento de las cargas de trabajo de ingeniería de datos

- Combinar procesamiento por lotes y de secuencias en una misma canalización
- Organizar el lago de datos en niveles de transformación de archivos
- Indexar el almacenamiento del lago de datos para la aceleración de consultas y cargas de trabajo

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Describir Azure Synapse Analytics
- Descripción de Azure Databricks
- Describir Azure Data Lake Storage
- Describir la arquitectura de Delta Lake
- Describir Azure Stream Analytics

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## Módulo 2: Ejecución de consultas interactivas con grupos de SQL sin servidor de Azure Synapse Analytics

En este módulo, los alumnos aprenderán a trabajar con archivos almacenados en el lago de datos y en orígenes de archivos externos mediante instrucciones T-SQL ejecutadas por un grupo de SQL sin servidor en Azure Synapse Analytics. Consultarán los archivos Parquet almacenados en un lago de datos, así como los archivos CSV almacenados en un almacén de datos externo. Luego, crearán grupos de seguridad de Azure Active Directory y exigirán el acceso a los archivos del lago de datos a través del control de acceso basado en roles (RBAC) y listas de control de acceso (ACL).

### Lecciones

- Exploración de las capacidades de los grupos de SQL sin servidor de Azure Synapse
- Consulta de datos en el lago mediante grupos de SQL sin servidor de Azure Synapse
- Creación de objetos de metadatos en grupos de SQL sin servidor de Azure Synapse
- Protección de datos y administración de usuarios en grupos de SQL sin servidor de Azure Synapse

### Laboratorio: Ejecución de consultas interactivas con grupos de SQL sin servidor

- Consultar datos de Parquet con grupos de SQL sin servidor
- Crear tablas externas para archivos Parquet y CSV
- Crear vistas con grupos de SQL sin servidor
- Proteger el acceso a los datos en un lago de datos cuando se usan grupos de SQL sin servidor
- Configurar la seguridad del lago de datos a través del control de acceso basado en roles (RBAC) y listas de control de acceso (ACL)

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Describir las capacidades de los grupos de SQL sin servidor de Azure Synapse
- Consulta de datos en el lago mediante grupos de SQL sin servidor de Azure Synapse
- Creación de objetos de metadatos en grupos de SQL sin servidor de Azure Synapse
- Protección de datos y administración de usuarios en grupos de SQL sin servidor de Azure Synapse

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## Módulo 3: Exploración y transformación de datos en Azure Databricks

En este módulo se enseña a usar varios métodos DataFrame de Apache Spark para explorar y transformar datos en Azure Databricks. Los alumnos aprenderán a usar métodos DataFrame estándar para explorar y transformar datos. También aprenderán a realizar tareas más avanzadas, como quitar datos duplicados, manipular valores de fecha y hora, cambiar el nombre de columnas y agregar datos.

### Lecciones

- Descripción de Azure Databricks
- Lectura y escritura de datos en Azure Databricks
- Trabajo con elementos DataFrame en Azure Databricks
- Trabajo con métodos avanzados de DataFrame en Azure Databricks

### Laboratorio: Realización de exploraciones y transformaciones de datos en Azure Databricks

- Usar DataFrames en Azure Databricks para explorar y filtrar datos
- Almacenar DataFrames en caché para realizar consultas más rápidas posteriormente
- Eliminación de datos duplicados
- Manipular valores de fecha y hora
- Quitar columnas de DataFrame y cambiarlas de nombre
- Agregar datos almacenados en un DataFrame

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Descripción de Azure Databricks
- Lectura y escritura de datos en Azure Databricks
- Trabajo con elementos DataFrame en Azure Databricks
- Trabajo con métodos avanzados de DataFrame en Azure Databricks

## Módulo 4: Exploración, transformación y carga de datos en almacenamientos de datos con Apache Spark

En este módulo se enseña a explorar los datos almacenados en un lago de datos, a transformar los datos y a cargarlos en un almacén de datos relacional. Los alumnos explorarán archivos Parquet y JSON y usarán técnicas para consultar y transformar archivos JSON con estructuras jerárquicas. Luego, usarán Apache Spark para cargar datos en el almacenamiento de datos y unirán datos de Parquet en el lago de datos con datos del grupo de SQL dedicado.

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## Lecciones

- Definición de la ingeniería de macrodatos con Apache Spark en Azure Synapse Analytics
- Ingesta de datos con cuadernos de Apache Spark en Azure Synapse Analytics.
- Transformación de datos con objetos DataFrame de grupos de Apache Spark de Azure Synapse Analytics
- Integración de grupos de SQL y Apache Spark en Azure Synapse Analytics

## Laboratorio: Exploración, transformación y carga de datos en almacenamientos de datos con Apache Spark

- Realizar exploraciones de datos en Synapse Studio
- Ingerir datos con cuadernos de Spark en Azure Synapse Analytics
- Transformar datos con DataFrame de grupos de Spark de Azure Synapse Analytics
- Integrar grupos de SQL y Spark en Azure Synapse Analytics

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Describir la ingeniería de macrodatos con Apache Spark en Azure Synapse Analytics
- Ingesta de datos con cuadernos de Apache Spark en Azure Synapse Analytics
- Transformación de datos con objetos DataFrame de grupos de Apache Spark de Azure Synapse Analytics
- Integración de grupos de SQL y Apache Spark en Azure Synapse Analytics

## Módulo 5: Ingesta y carga de datos en almacenamientos de datos

En este módulo se enseña a los alumnos a ingerir datos en el almacenamiento de datos mediante scripts de T-SQL y canalizaciones de integración de Synapse Analytics. Los alumnos aprenderán a cargar datos en grupos de SQL dedicados de Synapse con PolyBase y COPY mediante T-SQL. También aprenderán a usar la administración de cargas de trabajo junto con una actividad de copia en una canalización de Azure Synapse para la ingesta de datos a escala de petabytes.

## Lecciones

- Uso de procedimientos recomendados para la carga de datos en Azure Synapse Analytics
- Ingesta a escala de petabytes con Azure Data Factory

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## Laboratorio: Ingesta y carga de datos en almacenamientos de datos

- Realizar ingestas a escala de petabytes con canalizaciones de Azure Synapse
- Importar datos con PolyBase y COPY mediante T-SQL
- Uso de procedimientos recomendados para la carga de datos en Azure Synapse Analytics

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Uso de procedimientos recomendados para la carga de datos en Azure Synapse Analytics
- Ingesta a escala de petabytes con Azure Data Factory

## Módulo 6: Transformación de datos con canalizaciones de Azure Data Factory o Azure Synapse

En este módulo se enseña a los alumnos a crear canalizaciones de integración de datos para ingerir desde varios orígenes de datos, transformar datos mediante flujos de datos de asignación y realizar movimientos de datos en uno o varios receptores de datos.

### Lecciones

- Integración de datos con Azure Data Factory o canalización de Azure Synapse
- Realización de transformaciones sin código y a escala con canalizaciones de Azure Data Factory o Azure Synapse

## Laboratorio: Transformación de datos con canalizaciones de Azure Data Factory o Azure Synapse

- Ejecutar transformaciones sin código y a escala con canalizaciones de Azure Synapse
- Crear una canalización de datos para importar archivos CSV con formato deficiente
- Crear flujos de datos de asignación

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Realizar integraciones de datos con Azure Data Factory
- Realización de transformaciones sin código y a escala con Azure Data Factory

## CONTÁCTENOS

## Módulo 7: Organización de movimientos y transformaciones de datos en canalizaciones de Azure Synapse

En este módulo aprenderemos a crear servicios vinculados y a organizar el movimiento y la transformación de datos mediante cuadernos en canalizaciones de Azure Synapse.

### Lecciones

- Orquestación de movimientos y transformaciones de datos en Azure Data Factory

### Laboratorio: Organización de movimientos y transformaciones de datos en canalizaciones de Azure Synapse

- Integrar datos de Notebooks con canalizaciones de Azure Data Factory o Azure Synapse

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Organizar movimientos y transformaciones de datos en canalizaciones de Azure Synapse

## Módulo 8: Seguridad integral con Azure Synapse Analytics

En este módulo, los alumnos aprenderán a proteger un área de trabajo de Synapse Analytics y su infraestructura de apoyo. Analizarán el administrador de SQL Active Directory, administrarán reglas de firewall de IP, administrarán secretos con Azure Key Vault y accederán a esos secretos a través de un servicio vinculado de Key Vault y actividades de canalización. También aprenderán a implementar seguridad en el nivel de columna y en el nivel de fila y el enmascaramiento dinámico de datos al usar grupos de SQL dedicados.

### Lecciones

- Creación de un almacenamiento de datos en Azure Synapse Analytics
- Configuración y administración de secretos en Azure Key Vault
- Implementación de controles de cumplimiento para datos confidenciales

### Laboratorio: Seguridad integral con Azure Synapse Analytics

- Proteger la infraestructura tras Azure Synapse Analytics
- Proteger el área de trabajo y los servicios administrados de Azure Synapse Analytics
- Proteger los datos del área de trabajo de Azure Synapse Analytics

## CONTÁCTENOS



Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Creación de un almacenamiento de datos en Azure Synapse Analytics
- Configuración y administración de secretos en Azure Key Vault
- Implementación de controles de cumplimiento para datos confidenciales

## Módulo 9: Soporte de procesamientos analíticos transaccionales híbridos con Azure Synapse Link

En este módulo, los alumnos aprenderán cómo Azure Synapse Link permite una conectividad sin fisuras entre una cuenta de Azure Cosmos DB y un área de trabajo de Synapse. Los alumnos verán cómo habilitar y configurar Synapse Link y, tras ello, cómo consultar el almacén analítico de Azure Cosmos DB mediante Apache Spark y SQL sin servidor.

### Lecciones

- Diseño del procesamiento analítico y transaccional híbrido mediante Azure Synapse Analytics
- Configuración de Azure Synapse Link con Azure Cosmos DB
- Consulta de Azure Cosmos DB con grupos de Apache Spark
- Consulta de Azure Cosmos DB con grupos de SQL sin servidor

### Laboratorio: Soporte de procesamientos analíticos transaccionales híbridos con Azure Synapse Link

- Configuración de Azure Synapse Link con Azure Cosmos DB
- Consultar Azure Cosmos DB con Apache Spark para Synapse Analytics
- Consultar Azure Cosmos DB con grupos de SQL sin servidor para Azure Synapse Analytics

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Diseño del procesamiento analítico y transaccional híbrido mediante Azure Synapse Analytics
- Configuración de Azure Synapse Link con Azure Cosmos DB
- Consulta de Azure Cosmos DB con Apache Spark para Azure Synapse Analytics
- Consultar Azure Cosmos DB con SQL sin servidor para Azure Synapse Analytics

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## Módulo 10: Procesamiento de secuencias en tiempo real con Stream Analytics

En este módulo, los alumnos aprenderán a procesar datos de secuencias con Azure Stream Analytics. Ingerirán datos de telemetría de vehículos en Event Hubs y, tras ello, los procesarán en tiempo real mediante varias funciones basadas en ventana en Azure Stream Analytics. Enviarán los datos a Azure Synapse Analytics. Por último, los alumnos aprenderán a escalar el trabajo de Stream Analytics para aumentar el rendimiento.

### Lecciones

- Habilitación de mensajería confiable para aplicaciones de macrodatos con Azure Event Hubs
- Trabajo con secuencias de datos mediante Azure Stream Analytics
- Ingesta de flujos de datos con Azure Stream Analytics

### Laboratorio: Procesamiento de secuencias en tiempo real con Stream Analytics

- Usar Stream Analytics para procesar datos en tiempo real desde Event Hubs
- Usar funciones basadas en ventana de Stream Analytics para crear agregados y enviarlos a Synapse Analytics
- Escalar trabajos de Azure Stream Analytics para aumentar el rendimiento a través de la creación de particiones
- Volver a particionar la entrada de secuencias para optimizar la paralelización

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Habilitación de mensajería confiable para aplicaciones de macrodatos con Azure Event Hubs
- Trabajo con secuencias de datos mediante Azure Stream Analytics
- Ingesta de flujos de datos con Azure Stream Analytics

## CONTÁCTENOS

+57 316 3956090

contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)

## Módulo 11: Creación de una solución de procesamiento de secuencias con Event Hubs y Azure Databricks

En este módulo, los alumnos aprenderán a ingerir y procesar datos de secuencias a escala con Event Hubs y streaming estructurado de Spark en Azure Databricks. Los alumnos conocerán los usos y las características clave del streaming estructurado. Implementarán ventanas deslizantes para agregar fragmentos de datos y aplicarán marcas de agua para quitar datos obsoletos. Por último, los alumnos se conectarán a Event Hubs para leer y escribir secuencias.

### Lecciones

- Procesamiento de datos de streaming con Structured Streaming de Azure Databricks

### Laboratorio: Creación de una solución de procesamiento de secuencias con Event Hubs y Azure Databricks

- Analizar los usos y características clave del streaming estructurado.
- Transmitir datos de un archivo y escribirlos en un sistema de archivos distribuido
- Usar ventanas deslizantes para agregar fragmentos de datos en lugar de todos los datos
- Aplicar marcas de agua para quitar datos obsoletos
- Conectarse a flujos de lectura y escritura de Event Hubs

Después de completar este módulo, los alumnos podrán hacer lo siguiente:

- Procesamiento de datos de streaming con Structured Streaming de Azure Databricks

## CONTÁCTENOS

+57 316 3956090

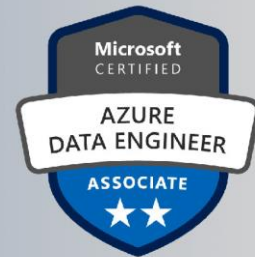
contactenos@siv.com.co

 [www.siv.com.co](http://www.siv.com.co)

+57 315 2653920

comercial@siv.com.co

 [www.dbcloud.co](http://www.dbcloud.co)



## ***DESCRIPCION CAPACITACION***

---

### **Duración de la Capacitación**

La capacitación tiene una intensidad de 32 horas.

### **Fechas y Horario Capacitación**

La capacitación en horario nocturno de 6:30 P.M. A 9:30 P.M. hora de Colombia 3 veces por semana.

### **Plataforma Capacitación**

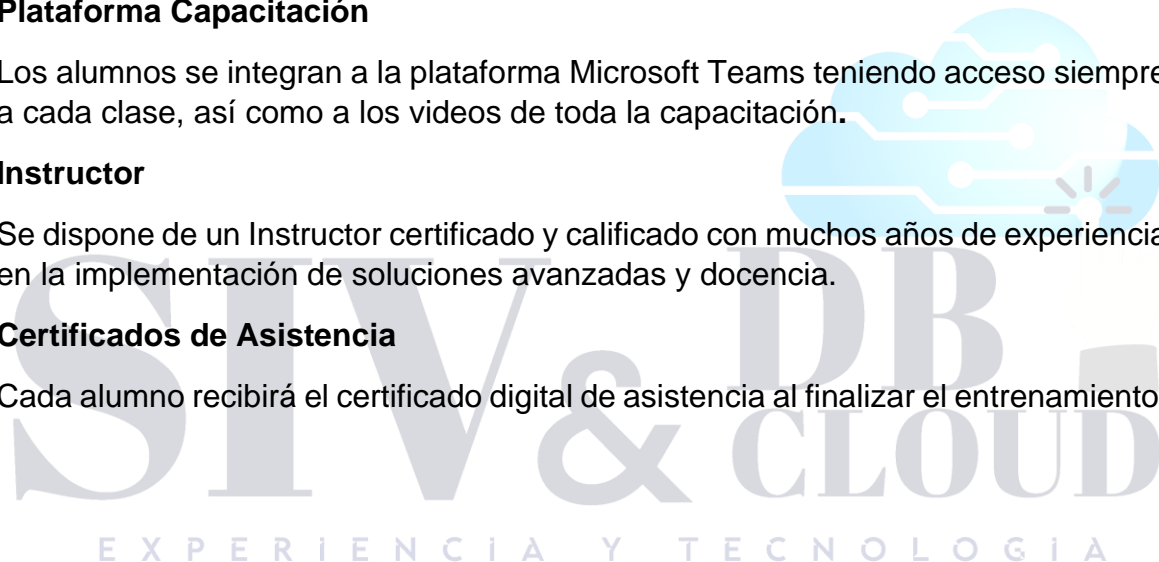
Los alumnos se integran a la plataforma Microsoft Teams teniendo acceso siempre a cada clase, así como a los videos de toda la capacitación.

### **Instructor**

Se dispone de un Instructor certificado y calificado con muchos años de experiencia en la implementación de soluciones avanzadas y docencia.

### **Certificados de Asistencia**

Cada alumno recibirá el certificado digital de asistencia al finalizar el entrenamiento.



## **CONTÁCTENOS**

 +57 316 3956090

 [contactenos@siv.com.co](mailto:contactenos@siv.com.co)

 [www.siv.com.co](http://www.siv.com.co)

 +57 315 2653920

 [comercial@siv.com.co](mailto:comercial@siv.com.co)

 [www.dbcloud.co](http://www.dbcloud.co)