

> Especialización en

AZURE DATA ENGINEERING



FORMACIÓN PARA LOS NUEVOS LÍDERES DE DATA & AI

DMC INSTITUTE, según el Estudio de Adopción de la Ciencia de Datos y Brecha de Talento 2025 de Apoyo Consultoría, es la institución no universitaria

Nº1

más valorada por

**Chief Data Officers y
empleadores del Perú**

para la formación en

CIENCIA DE DATOS

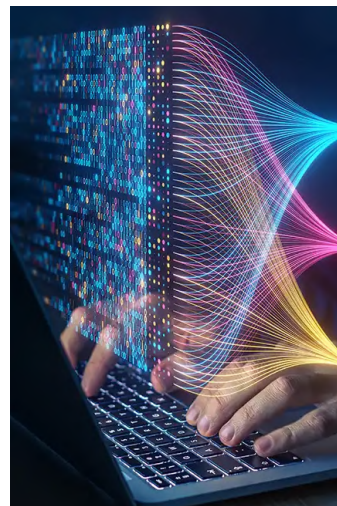
Fuente: Estudio de Adopción de la
Ciencia de Datos y Brecha de Talento 2025.
Apoyo Consultoría



¿Por qué llevar esta especialización?

¿Sabías que Azure ofrece un ecosistema completo que permite construir soluciones modernas de datos con servicios como Data Factory, Synapse Analytics, Databricks y Power BI, y todo en una plataforma unificada?

DMC Institute ha diseñado la Especialización Azure Data Engineering que te permitirá dominar el diseño y despliegue de pipelines de datos y almacenes modernos de información como datalakes y lakehouses en la nube de Microsoft. Aprenderás a construir soluciones escalables, seguras y automatizadas, aprovechando la elasticidad de Azure para manejar desde cargas pequeñas hasta petabytes de datos.



Sobre esta Especialización

14

sesiones

56

horas
académicas

34

talleres
prácticos

¿Cómo impulsamos tu carrera?

- Sesiones 80% enfocadas en la práctica.
- Enfoque en **Casos Aplicados a Negocio**, enfrentando los retos del mercado.
- Énfasis en **habilidades técnicas**.
- **Mentoría especializada** con docentes expertos.
- Acompañamiento **constante**.



Objetivos de la Especialización

OBJETIVO GENERAL:

- Utiliza los servicios de la plataforma de Ms. Azure para el diseño e implementación de Lakehouses adaptados a diversos entornos y volúmenes de procesamiento, contemplando fuentes batch y streaming de diferentes fuentes y formatos.

Requisitos

- Comprensión de bases de datos relacionales y sus principales conceptos: tabla, relación, cardinalidad.
- Conocimiento de un lenguaje de programación.
- Mínimo un año de experiencia trabajando en áreas de datos, o elaborando reportes e informes.
- Contar con una laptop o computadora de escritorio con disponibilidad de micrófono y cámara web.
- Tener instalado los softwares y herramientas señalados en la sección Contenidos.
- Contar con una cuenta de Microsoft Azure.

Dirigido a

- **Ingenieros de datos Jr.**, que actualmente busquen:
 - Especializarse en el manejo de la plataforma Microsoft Azure para la implementación de sus soluciones de datos.
- **Ingenieros de datos expertos en sistemas on-premise o en otras plataformas cloud**, que estén buscando:
 - Migrar sus soluciones a Microsoft Azure, o crear entornos híbridos con ella.
- **Analistas y científicos de datos**, que estén buscando:
 - Incorporar a sus proyectos de visualización y/o de machine learning un motor de datos robusto basado en la nube de Microsoft Azure.
 - Complementar o enrumbar su carrera hacia la ingeniería de datos en la nube.
- **Profesionales de las áreas de sistemas, informática, desarrollo de software y afines**, que estén buscando:
 - Iniciar y especializarse en la carrera de ingeniería de datos on cloud, de rápido crecimiento y alta demanda laboral.



Malla Curricular

I. Azure Fundamentals

1. Cloud Fundamentals

- Cloud Computing. Definición y ventajas. Diferencias respecto a sistemas on-premise.
- Clasificación clásica de servicios: IaaS, PaaS, SaaS.
- Evolución de los servicios cloud. Everything as a Service (XaaS).
- Servicios dedicados vs Serverless. Diferencias principales, casos de uso.
- Modelos de costo generales. Por almacenamiento, por capacidad de cómputo, por flujo de datos, entre otros.

2. Azure Fundamentals

- **Taller:** Exploración de los componentes principales de la consola de Azure.
- **Taller:** Introducción a los servicios de cómputo, creación y testeo de una máquina virtual.
- **Taller:** Introducción a los servicios de almacenamiento, creación y testeo de Blob y ADLS Gen2.
- **Taller:** Introducción a los servicios de bases de datos, creación y testeo de bases de datos relacionales.
- **Taller:** Introducción a los servicios de seguridad. Configuración y uso de Azure AD y RBAC.
- **Taller:** Introducción a los servicios de monitoreo. Recomendaciones para minimizar el billing.

II. Azure Lakehouse Fundamentals (Low-code)

3. Introducción a Data Warehousing

- Data Warehouse vs. BD relacional. Repaso de motores DW disponibles.
- Arquitecturas modernas para Warehouses.
- Modelos tabulares y multidimensionales. Aplicaciones en analytics y data science.

4. Proyecto 1. Warehouse básico con Dedicated SQL Pool (Synapse)

- **Taller:** Despliegue de un SQL Pool y división lógica en dos capas (Raw-Curated).
- **Taller:** Implementación de un ETL para ingesta de archivos (CSV) desde un bucket, empleando Synapse Studio y lenguaje SQL.
- **Taller:** Implementación de un ETL para ingesta incremental desde un motor de base de datos, empleando Synapse Studio y lenguaje SQL.
- **Taller:** Implementación de ETLs sencillos para procesamiento (Raw a Curated) empleando Synapse Studio.
- **Taller:** Automatización de los ETLs por calendario empleando Trigger Schedule.
- **Taller:** Test de conexión desde Power BI Desktop.

5. Proyecto 2. Lakehouse básico sobre un bucket ADLS Gen2 con ETLs visuales

- **Taller:** Despliegue de un Datalake empleando ADLS Gen2 y división en dos capas lógicas (Raw-Curated).
- **Taller:** Implementación de un ETL para ingesta de archivos (CSV) desde otro bucket, empleando la interfaz visual de Data Factory (Data Flow).
- **Taller:** Implementación de un ETL para ingesta incremental desde un motor de base de datos, empleando la interfaz visual de Data Factory (Data Flow) y basado en Timestamp.
- **Taller:** Implementación de un ETL para procesamiento (Raw a Curated) empleando la interfaz visual de Data Factory (Data Flow).
- **Taller:** Automatización de los ETLs por calendario empleando Trigger Schedule de Data Factory.
- **Taller:** Exposición del datalake empleando SQL Pool Serverless. Test de conexión desde Power BI Desktop.

6. Buenas prácticas y optimización

- Revisión conceptual de arquitecturas con soporte Push-down para optimización de respuesta.
- Revisión conceptual sobre escalamiento de Integration Runtime (IR) y paralelización de pipelines.
- Consideraciones para optimización de costos.

III. Azure Scalable Lakehouses

7. Introducción a Medallion y Lakehouses escalables

- Enfoques ETL vs. ELT para datalakes modernos.
- Arquitectura Medallion. Diseño arquitectónico y aplicaciones en Data Engineering.
- Formatos abiertos y exclusivos para Medallion. Iceberg, Hudi, Delta. Ventajas y aplicaciones de cada uno.
- Storage vs. Processing escalables. Herramientas y servicios disponibles en Azure.

8. Proyecto 1. Lakehouse Medallion con fuentes streaming directas y motor de procesamiento nativo (HDInsight) con PySpark

Implementación del Datalake

- **Taller:** Despliegue de un Datalake empleando ADLS Gen2 y división lógica según Medallion.

Ingesta semi Realtime (desde origen exclusivo)

- **Taller:** Implementación de un bucket como repositorio de ingesta exclusivo.
- **Taller:** Implementación de un ETL basado en código Python para ingesta de archivos, empleando Custom Activity de Data Factory. Automatización del ETL por evento empleando Event Grid (sin filtrado de eventos).

Malla Curricular

Ingesta streaming directa

- **Taller:** Implementación de un ETL para captura streaming directa desde un motor de base de datos, empleando Azure Database CDC y Event Hubs.
- **Taller:** Implementación de un ETL para captura streaming directa desde una API, empleando Event Hubs y Azure Functions.

Flujo de datos y procesamiento

- **Taller:** Implementación de ETLs PySpark para flujo y procesamiento de datos entre las capas del Datalake, empleando HDInsight.
- **Taller:** Automatización de los ETLs por calendario y por eventos mediante servicios basados en Airflow, empleando Azure Managed Airflow.

Exposición del Datalake

- **Taller:** Exposición del datalake empleando SQL Pool Serverless. Test de conexión desde Power BI Desktop.

9. Proyecto 2. Lakehouse Medallion con fuentes streaming transformadas y procesamiento sobre Databricks con PySpark Ingesta semi Realtime (desde origen compartido)

- **Taller:** Implementación de un bucket como repositorio de ingesta compartido.
- **Taller:** Implementación de un ETL basado en código Python para ingesta de archivos, empleando Custom Activity de Data Factory. Automatización del ETL por evento empleando Event Grid (con filtrado de eventos).

Ingesta streaming con transformación

- **Taller:** Implementación de un ETL para captura streaming y transformación desde un motor de base de datos, empleando Azure Database CDC, Event Hubs y Stream Analytics.
- **Taller:** Implementación de un ETL para captura streaming y transformación desde una API, empleando Event Hubs y Stream Analytics.

Flujo de datos y procesamiento

- **Taller:** Montaje del datalake al entorno de Databricks.
- **Taller:** Implementación de ETLs PySpark para flujo y procesamiento de datos entre las capas del Datalake.
- **Taller:** Automatización de los ETLs por calendario y evento en Databricks.

Exposición del Datalake

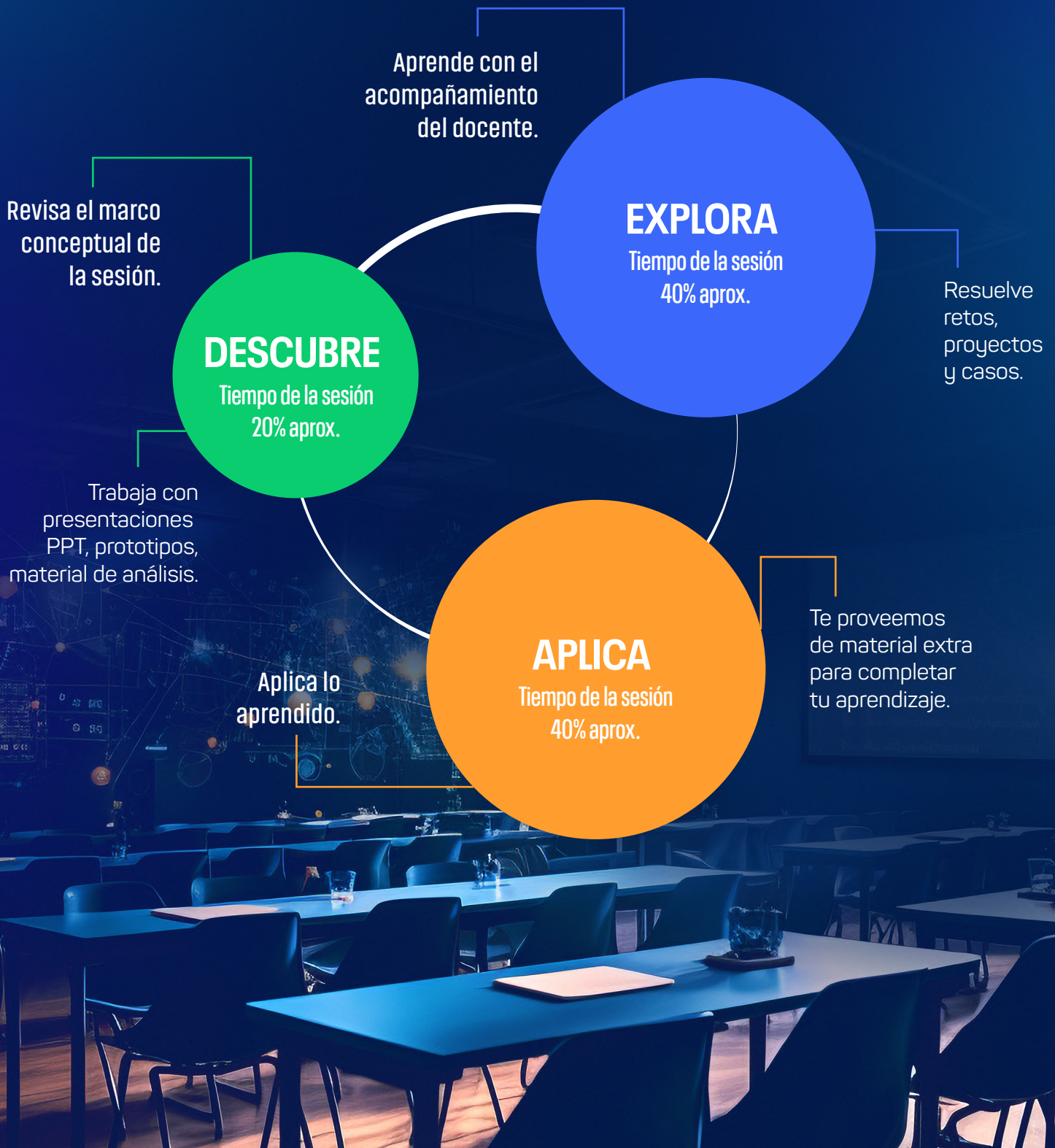
- **Taller:** Test de conexión desde Power BI Desktop.

10. Buenas prácticas y optimización

- Formatos de tabla recomendados para cada capa de Medallion.
- Datalake (ADLS Gen2) vs. Databricks vs. Synapse Serverless como motores de datos. Consideraciones y casos de uso.
- Revisión conceptual de técnicas de optimización en Databricks. Particionado, Z-Ordering, auto-optimize, Vacuum en Delta Lake.
- Consideraciones para optimización de costos.

METODOLOGÍA LEARNING AGILE

Prepárate para nuevos retos, aprende con nuestro método de primer nivel y lánzate a conquistar el mundo.



¿Qué certificado obtendrás?

- Certificado por aprobación de la Especialización en Azure Data Engineering, por un total de 56 horas académicas.



¿Por qué elegirnos?

-  Somos los primeros en Perú en apostar por el desarrollo de profesionales y empresas en data & analytics con más de 16 años de experiencia.
-  Las empresas worldclass de Latam confían en nosotros para acompañarlas en su transformación hacia el enfoque data driven.
-  Nuestros docentes son destacados expertos en data & analytics que lideran equipos de alto rendimiento en las empresas más grandes de Latam.
-  Nuestra metodología "Aprende haciendo" ha logrado que nuestra comunidad de +25K profesionales en todo Latam mejoren su situación laboral.
-  Tenemos el portafolio más completo con +150 capacitaciones sincrónicas y asincrónicas que se ajustan a diferentes perfiles y niveles de conocimiento.





Visita nuestra web

www.dmc.pe

Síguenos en:

