

> Especialización en

AWS DATA ENGINEERING



¿Por qué llevar esta especialización?

¿Sabías que AWS ofrece un ecosistema robusto y probado para ingeniería de datos, con servicios como Amazon S3, Glue, Redshift, Athena y Kinesis, que permiten construir soluciones distribuidas y adaptables a cualquier necesidad de negocio?

Por ello, DMC INSTITUTE ha diseñado la Especialización AWS Data Engineering que te permitirá aprender a diseñar, implementar y escalar arquitecturas de datos modernas utilizando lo mejor de la nube de Amazon. Aprovecharás las capacidades serverless, el procesamiento en tiempo real (streaming) y el almacenamiento masivo para transformar datos en valor, sin preocuparte por la infraestructura subyacente.



Sobre esta Especialización

14

sesiones

56

horas
académicas

32

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 Amazon Web Services (AWS) 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 sistema operativo Windows y disponibilidad de micrófono y cámara web.
- Contar con una cuenta de Amazon Web Services (AWS).

Dirigido a

- **Ingenieros de datos Jr.**, que actualmente busquen:
 - Especializarse en el manejo de la plataforma de Amazon Web Services (AWS) 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 la nube de Amazon, 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 AWS.
 - 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. AWS 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. AWS Fundamentals

- **Taller:** Exploración de los componentes principales de la consola de AWS.
- **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 un S3.
- **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 IAM y KMS.
- **Taller:** Introducción a los servicios de monitoreo. Recomendaciones para minimizar el billing.

II. AWS 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 sobre Amazon Redshift con procesamiento SQL

- **Taller:** Despliegue de un clúster Redshift y división lógica en dos capas (Raw a Curated).
- **Taller:** Implementación de un ETL para ingesta de archivos (CSV) desde un bucket, empleando lenguaje SQL (store procedure).
- **Taller:** Implementación de un ETL para ingesta incremental desde un motor de base de datos, empleando lenguaje SQL (store procedure).
- **Taller:** Implementación de ETLs sencillos para procesamiento (Raw a Curated) empleando lenguaje SQL (store procedure).
- **Taller:** Automatización de los ETLs por calendario empleando Event Bridge y Lambda.

5. Proyecto 2. Lakehouse básico sobre un bucket S3 con ETLs visuales

- **Taller:** Despliegue de un Datalake empleando un bucket S3 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 Glue Studio.
- **Taller:** Implementación de un ETL para ingesta incremental desde un motor de base de datos, empleando la interfaz visual de Glue Studio (basado en Timestamp).
- **Taller:** Implementación de un ETL sencillo de procesamiento (Raw a Curated) empleando la interfaz visual de Glue Studio.
- **Taller:** Automatización de los ETLs por calendario empleando Glue Workflows.
- **Taller:** Exposición del Datalake empleando Amazon Athena. Test de conexión desde QuickSight y Power BI Desktop.

6. Buenas prácticas y optimización

- Revisión conceptual de arquitecturas con soporte Push-down para optimización de respuesta.
- Consideraciones para optimización de costos.

III. AWS 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 de tabla para Medallion. Iceberg, Hudi, Delta. Ventajas y aplicaciones de cada uno.
- Storage vs. Processing escalables. Herramientas y servicios disponibles en AWS.

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

Implementación del Datalake

- **Taller:** Despliegue de un Datalake empleando un bucket S3 y división lógica según Medallion

Ingesta semi Realtime (desde origen exclusivo)

- **Taller:** Implementación de un bucket como repositorio de ingesta exclusivo (único fin).
- **Taller:** Implementación de un ETL basado en código Python (PySpark), empleando Custom Transform de Glue Studio. Automatización del ETL por evento empleando EventBridge (sin filtrado de eventos).

Malla Curricular

Ingesta streaming directa

- **Taller:** Implementación de un ETL para captura streaming directa e incremental desde un motor de base de datos, empleando AWS Database Migration Services.
- **Taller:** Implementación de un ETL para captura streaming directa desde una API, empleando Firehose y KDS.

Flujo de datos y procesamiento

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

Exposición del Datalake

- **Taller:** Exposición del Datalake empleando Amazon Athena. Test de conexión desde QuickSight y Power BI Desktop.

9. Proyecto 2. LLakehouse 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 Transform de Glue Studio. Automatización del ETL por evento empleando EventBridge (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 AWS Database Migration Services y Glue Streaming.

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 QuickSight y desde Power BI Desktop.

10. Buenas prácticas y optimización

- Formatos de tabla recomendados para cada capa de Medallion.
- Datalake (S3) vs. Databricks como motores independientes 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.

Nuestra Propuesta de Capacitación

Las metodologías que aplicamos



Desarrollo de competencias clave en el mundo de los datos

Analiza · Innova · Transforma



Aprendizaje Secuencial

- Descubre conocimiento de vanguardia
- Explora con la guía del experto
- Aplica lo aprendido



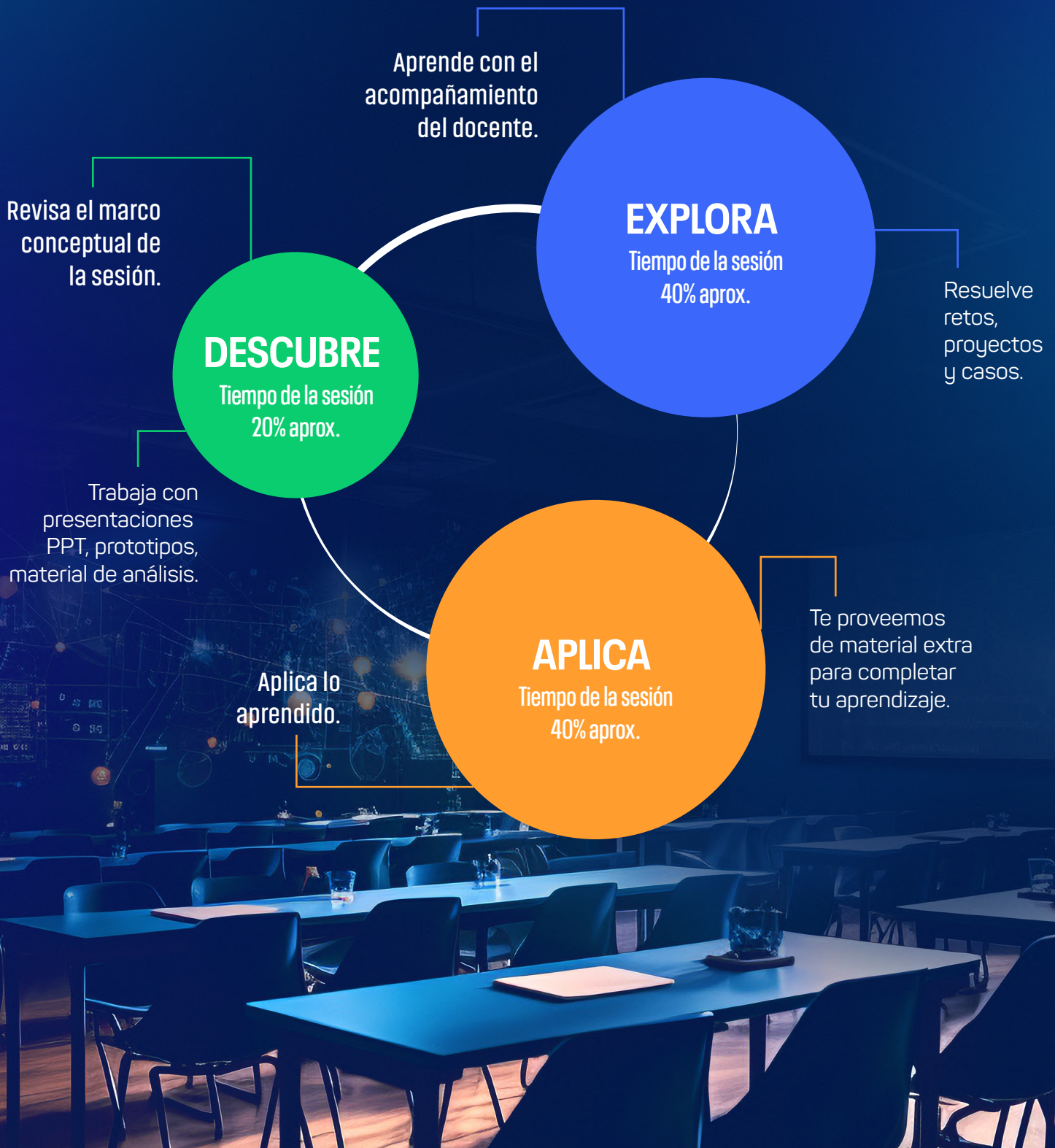
Aprendizaje basado en práctica (Learning by Doing)

- Resuelve retos
- Aprende en base a proyectos
- Analiza casos



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 AWS 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:

