

> Especialización en

BIG DATA



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



Presentación

Las empresas que vienen adoptando Big Data, requieren de profesionales que puedan afrontar y solucionar los retos asociados, en primer lugar mantener el ritmo de crecimiento de los datos y su almacenamiento eficiente, luego encontrar técnicas para su procesamiento y organización de tal modo que pueda realizarse análisis significativos para el negocio, y finalmente mantener y emplear tecnología tanto tradicional como nueva.

Por ello, DMC INSTITUTE presenta la "Especialización en Big Data", la cual te permitirá conocer sus fundamentos, la arquitectura tecnológica subyacente, y sobre todo las herramientas y los lenguajes disponibles para la manipulación de los datos, tanto para entornos On-premise como cloud.



Sobre esta Especialización

15

sesiones

60

horas
académicas

08

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:

- Aprende sobre los fundamentos de Big Data, la arquitectura tecnológica que la soporta, las herramientas y los lenguajes disponibles para la manipulación de los datos, tanto para entornos On-premise como cloud, y sobre todo su integración con soluciones de ingeniería de datos.

OBJETIVOS ESPECÍFICOS:

- Aprende sobre el almacenamiento distribuido y a utilizar diversas herramientas para la implementación de Data-Lakes y para el tratamiento de datos a gran escala en modos Batch y Real-Time, como son Apache Hive, Apache Spark, Databricks, y Apache Kafka.
- Aprende a utilizar los principales servicios de los proveedores cloud líderes en el mercado como Azure, GCP y AWS, para el diseño e implementación de ETL básicos.

Requisitos

- Sólidos conocimientos de programación, de preferencia con lenguaje Python.
- Contar con una laptop o computadora de escritorio con disponibilidad de micrófono y cámara web.
- Tener instalado los softwares y herramientas señaladas en la sección Contenidos.

Dirigido a

- Profesionales de ingeniería de sistemas, informática, desarrollo de software y similares interesados en incursionar en soluciones de ingeniería de datos e implementación de pipelines.
- Profesionales y consultores que laboren en áreas de inteligencia de negocios o ciencias de datos interesados en conocer tecnologías y herramientas que soporten y automaticen los flujos de datos para sus proyectos habituales.
- Profesionales y consultores que laboren en el área de ingeniería de datos interesados en actualizar su tool-set con herramientas y técnicas de vanguardia en este campo de los datos.



Malla Curricular

I. Big Data Processing

1. Introducción a Big Data

- Big Data. Definición, filosofía, las Vs.
- Big Data como marco de trabajo.
- Arquitectura conceptual.
- Componentes tecnológicos disponibles.
- Arquitectura moderna de datos.

2. Almacenamiento distribuido con Apache Hadoop

- Tecnologías Open-source para Big Data.
- Fundamentos de Apache Hadoop.
- Almacenamiento distribuido en HDFS.
- Taller: Procesamiento de datos con Apache Hive.
- Diferencias de Map Reduce vs Tez vs Apache Spark.
- Datalake. Definición y arquitectura (capas).
- Taller: Poblamiento de un Datalake con Apache Tez, Apache Hive y HDFS.

3. Procesamiento distribuido con Apache Spark

- Introducción a Spark.
- Funciones con PySpark.
- Extracción y Transformación de datos.
- Dataframes y RDDs.
- Funciones de Apache Spark.
- Tuning en Apache Spark

4. Ingeniería de datos con Databricks

- ¿Qué es Databricks?
- Databricks Community vs Azure Databricks.
- Conociendo la interfaz de Databricks.
- Data lakehouse vs Datalake.
- Iceberg vs Delta lake vs Apache Hudi.
- Taller: Creación de una Lakehouse con arquitectura Medallion con Delta Lake.

5. Real-time Fundamentals

- Real-time en datos. Definición, casos de aplicación, diferencias respecto a Batch.
- Arquitectura Publicador – Suscriptor.
- Tecnología para carga Real-Time en Big Data.
- Taller: Uso y configuración de Apache Kafka para carga de datos Real-Time.
- Apache Spark & Kafka setup environments.
- Taller: Integración de Apache Spark con Apache Kafka para procesamiento de datos Real-time.

II. Cloud Data Engineering

6. Fundamentos de Cloud Computing

- ¿Qué es computación en la nube?
- Conceptos de virtualización, Uso bajo demanda, despliegue escalable.
- Ventajas del cloud computing.
- Regiones y zonas de disponibilidad.
- Tipos de nubes.
- IAAS, PAAS y SAAS.

7. Introducción a la Ingeniería de datos con AWS

- Introducción a la Consola de AWS.
- Principales servicios de data en AWS. S3, EMR, Cloud Funtions, IAM, Redshift, Athena, entre otros.
- Arquitectura de datos en AWS.
- Taller: Diseño e implementación de un ETL básico con AWS.

8. Introducción a la Ingeniería de datos con Azure

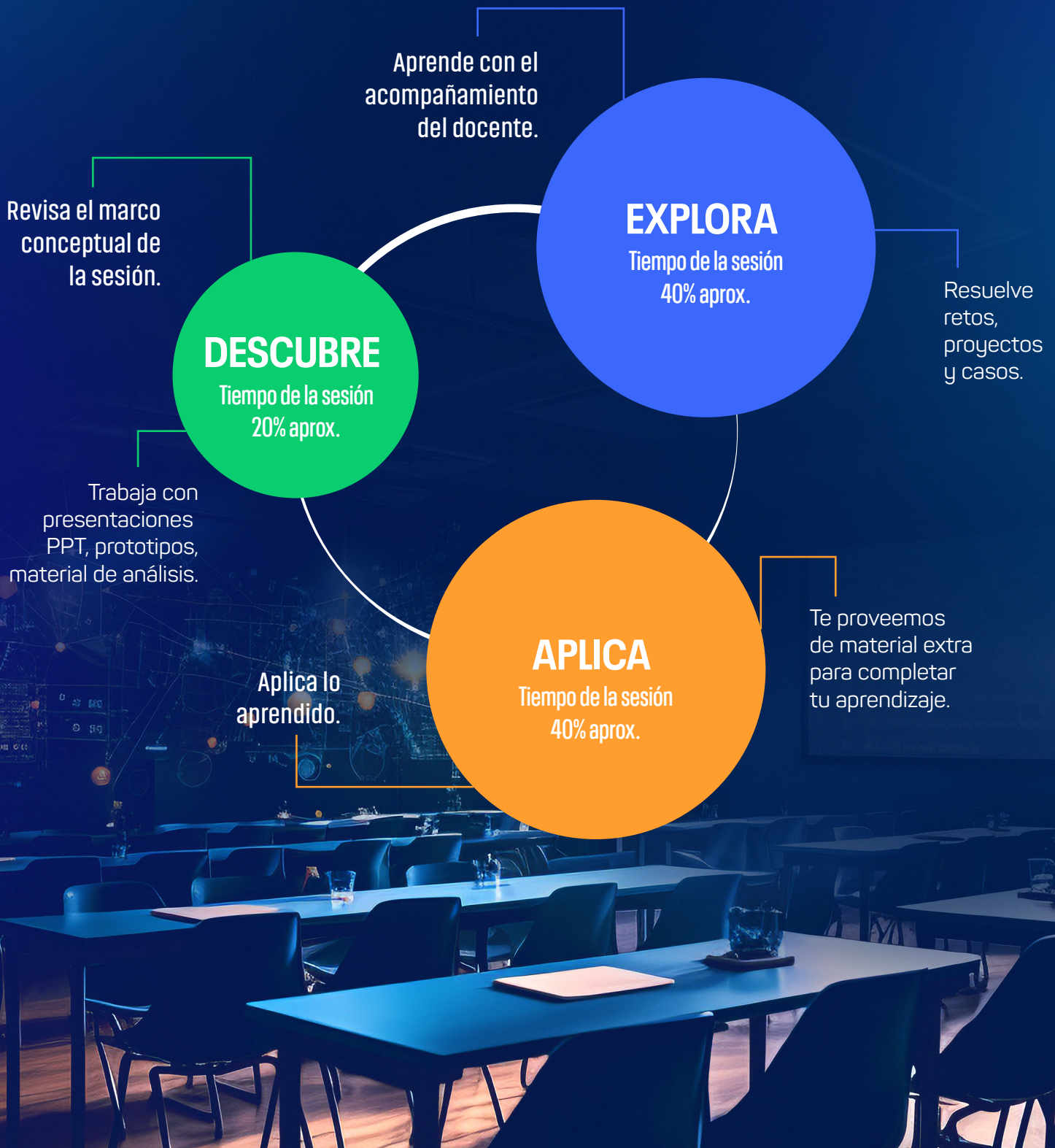
- Introducción a la Consola de Azure.
- Principales servicios de data en Azure. Blob Storage, Data factory, Databricks, Synapse, Event Hub, Azure SQL.
- Arquitectura de datos en AZURE.
- Taller: Diseño e implementación de un ETL básico con Azure.

9. Introducción a la Ingeniería de datos con GCP

- Introducción a la Consola de Azure.
- Servicios principales servicios de data en GCP. Cloud functions, Cloud Storage, Bigquery, Dataproc, data fusion, composer.
- Arquitectura de datos en GCP.
- Taller: Diseño e implementación de un ETL básico con GCP.

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 Big Data, por un total de 60 horas académicas.



¿Qué certificado obtendrás?

Al finalizar la capacitación, tendrás la oportunidad de acceder al **Big Data Professional Certification (CAIPC)** de CertiProf con un descuento especial.

Costo del examen 150 dólares	67.33%	Precio con descuento 49 dólares
--	---------------	---



IMPORTANTE:

- **No es una certificación automática de CertiProf**

Completar la capacitación no garantiza automáticamente el certificado de CertiProf, es un paso previo, pero no sustituye el proceso oficial de certificación.

- **Aprueba la capacitación y activa tu beneficio**

El primer paso para acceder al beneficio es aprobar la capacitación. ¡Superarlo te abre la puerta!

- **Tú eliges si aplicar o no**

El beneficio es opcional. Solo tú puedes decidir aprovecharlo

- **Más de una oportunidad para alcanzar tu meta**

Con el beneficio, tendrás hasta 2 intentos para rendir el examen y asegurar que tu esfuerzo sea recompensado con la certificación.

¿Por qué elegirnos?

+16

Más de 16 años
de experiencia.

+300

Más de 300 empresas
asesoradas en Perú,
Ecuador y Bolivia.

35k

35 mil profesionales
capacitados en más de 20
países de América Latina.



Propuesta integra
en formación en
Data & AI.

+150

Más de 150 docentes
expertos de Latinoamérica,
España y Estados Unidos.



Comunidad más grande en Data & AI
con beneficios exclusivos:
Networking, empleabilidad,
habilidades blandas.



Excelente nivel de
servicio.





Visita nuestra web

www.dmc.pe

Síguenos en:

