

> Especialización en
BIG DATA

60 horas académicas

100% Online **en vivo**

Certificación :



- **Por DMC INSTITUTE:** Por haber aprobado la Especialización en Big Data
- **Por CertiProf (Opcional):** Big Data Professional Certificate - BDCP

Presentación

Big Data se refiere tanto al gran volumen de datos como a las tecnologías necesarias para gestionarlos. Las organizaciones actuales enfrentan desafíos como el crecimiento acelerado de la información, su almacenamiento eficiente y su análisis en tiempo real.

La Especialización en Big Data te permitirá comprender sus fundamentos, conocer la arquitectura tecnológica que lo soporta y dominar herramientas y lenguajes clave para el manejo de datos, tanto en entornos on-premise como en la nube.



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:

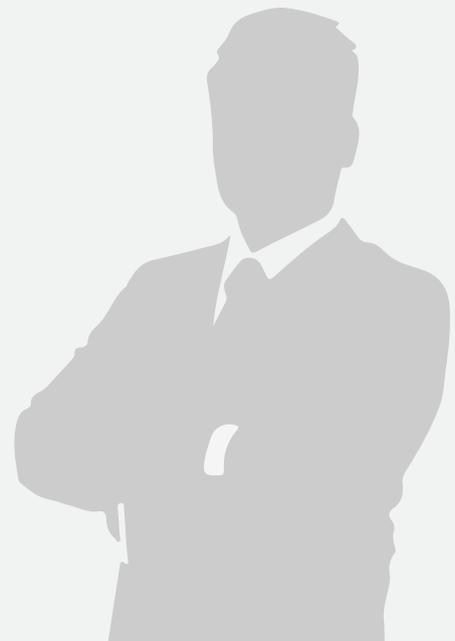
- Aprende sobre el almacenamiento distribuido y a utilizar diversas herramientas para la implementación de Data-Lakes y para el tratamiento de datos a gran escala en modos Batch y Real-Time, como son Apache Hive, Apache Spark, Databricks, y Apache Kafka.
- Aprende a utilizar los principales servicios 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.

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.



Malla Curricular

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.

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.

Malla Curricular

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.

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



¿Qué certificado obtendrás?

- Certificado por aprobación de la Especialización en Deep Learning & IA, por un total de 56 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.



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?



Somos los primeros en Perú en apostar por el desarrollo de profesionales y empresas en data & analytics con más de 15 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.



DMC | INSTITUTE
*Analiza, aplica, **transforma***

Visita nuestra web

www.dmc.pe

Síguenos en:     