

> Especialización en  
**BIG DATA**

**64 horas** académicas

100% Online **en vivo**

**Certificación :**

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



## ¿Por qué llevar esta especialización?

El término Big Data hace alusión tanto a la magnitud de los datos que representa como a la tecnología que la soporta. La literatura la describe a partir de un conjunto de cualidades denominadas las "V's del Big Data", y, aunque es variable la cantidad de ellas respecto a cada autor, tres son transversales: Volumen, Velocidad y Variedad. En esa línea, 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, y finalmente mantener y emplear tecnología tanto tradicional como nueva.

Por ello, DMC Perú 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.



## 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.
- Tunning 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.

# Metodología DMC



### Aprende Haciendo

Desarrolla casos con datos reales, incluso puedes proponer casos de tu propio sector.



### Clases en Vivo

El 100% de las clases que se desarrollan en el programa son en vivo.



### Asesoría Académica

Resuelve tus dudas con el asistente académico en línea.



### Plataforma E-learning

Accede en cualquier momento a materiales complementarios: videos, clases grabadas, etc.

# ¿Qué certificado obtendrás?

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



# ¿Qué certificado obtendrás?

- Al completar la Especialización de Big Data tendrás la opción de rendir el Big Data Professional Certificate (BDPC) emitido por Certiprof por un costo de 50 dólares. El costo real del examen de Certiprof es de 150 dólares.



# ¿Por qué elegirnos?



Somos los primeros en Perú en apostar por el desarrollo de profesionales y empresars en data & analytoics 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.





[www.dmc.pe](http://www.dmc.pe)