

Online / 

ESPECIALIZACIÓN EN

BIG DATA

BROCHURE 2024

#CienciaDeDatosDMC
#LíderesEnAnalítica



PRESENTACIÓN

El término Big Data, tan popularizado en los últimos años, 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 de tal modo que pueda realizarse análisis significativos para el negocio, 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.

-
-  **Duración:** 60 horas académicas
 -  **Certificación:** Por haber aprobado la Especialización en Big Data
 -  **Pre-requisito:** Sólidos conocimientos de programación, de preferencia con lenguaje Python.





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.



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, o desempeñen tales funciones, interesados en actualizar su tool-set con herramientas y técnicas de vanguardia en este campo de los datos.



CARACTERÍSTICAS

Clases en Vivo

El 100% de las clases que se desarrollan en la especialización son en vivo.

Asesoría Académica

Resuelve tus dudas con el asistente académico en línea.

Plataforma E-Learning

Accede en cualquier momento a materiales complementarios: lecturas, videos, tutoriales, clases grabadas y más.

Aprende haciendo

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

Certificación

Con una nota mínima de 14 sobre 20 y una asistencia mínima del 80%

Soporte técnico

Asistencia técnica permanente y acceso a máquinas virtuales de ser necesario.



MALLA CURRICULAR



Big Data Processing

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.

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.

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.

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.

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.

MALLA CURRICULAR



Cloud Data Engineering

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.

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.

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.

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.

Buscamos liderar la transformación de las empresas.



+14 Años de Experiencia

Desde el año 2009 capacitamos con técnicas de análisis de datos a profesionales de diferentes empresas y sectores.



+18K Profesionales Capacitados

Nuestros alumnos inscritos pertenecen a las mejores compañías del medio y amplifican con nosotros su red de contactos especializada.



+100 Empresas Asesoradas

Las empresas top del mercado buscan nuestra asesoría y les brindamos soluciones analíticas ad hoc.



+80 Líderes en Analítica

Nuestra plana docente ocupa los cargos más importantes en las áreas analíticas de todos los sectores.



+35 Capacitaciones Especializadas

Contamos con una variedad de líneas temáticas y niveles de especialización.



+5 Big Data Analytics Summit

Organizamos el evento más grande del Perú, con los mejores ponentes internacionales.



Formando profesionales mediante la analítica de los datos.

יחיד

