

Diploma

DATA ENGINEER



Presentación

El avance tecnológico ha promovido el crecimiento exponencial de los datos almacenados por las empresas, que es lo que hoy se denomina Big Data, que a su vez ha motivado el surgimiento de nuevas técnicas y herramientas de recolección, organización y almacenamiento para tales volúmenes de información. Ahora, el reto está en proveer soluciones que transporten y democratizen la data en toda la organización, y constituyan un soporte confiable.

Por ello DMC Perú, presenta el **Diploma Data Engineer** donde aprenderás el diseño e implementación de soluciones ETL desde diversas fuentes hacia repositorios como Data-Warehouses, Data-Lakes, entre otros; empleando lenguajes de programación líderes en el mercado como Python y Transact-SQL, contemplando en el proceso el paradigma y las herramientas para datos masivos, y complementos importantes como Web Scraping, Real-Time Data, DataOps, entre otros.



Información académica

 **Horas:** 168 horas académicas

 **Modalidad:** Online en vivo



Certificación triple

Por DMC:

- Especialista en Big Data
- Por haber aprobado el Diploma Data Engineer

Por Certiprof:

- Big Data Professional Certificate BDPC

¿A quién está dirigido?



Ing. de sistemas, informática, desarrollo de software o afines

Profesionales de estos campos de estudio que laboren en áreas relacionadas a datos que busquen:

- Incursionar y especializarse en la ingeniería de datos.
- Orientar sus conocimientos de programación y de bases de datos hacia la implementación de soluciones de ETL.

Expertos en bases de datos Y DBAs

Profesionales que se desempeñen en estos puestos y busquen:

- Acondicionar sus sistemas de almacenamiento para el manejo de datos masivos y Big Data.
- Migrar y replicar sus sistemas de almacenamiento a entornos cloud, garantizando flujos de datos confiables y totalmente automatizados.

Requisitos

- Experiencia profesional mayor a un año en actividades relacionadas a la manipulación y/o procesamiento de datos.
- Preferentemente PC de escritorio con disponibilidad de micrófono y cámara web.
- Tener instalado los softwares y herramientas señalados en la sección Contenidos.

¿Qué aprenderás en el Diploma?

- A diseñar e implementar ETL con los lenguajes de programación Python y T-SQL para la ingesta, transformación y carga de datos entre bases de datos relacionales.
- A implementar soluciones para el almacenamiento y tratamiento de datos masivos (Big Data) mediante el empleo individual y concurrente de herramientas open-source como Apache Spark, Apache Hive, Apache Tez, Apache Kafka, entre otras.
- A seleccionar e implementar soluciones complementarias a los flujos de ingeniería de datos a fin de garantizar su control, automatización y eficiencia.
- A integrar todas las soluciones y herramientas mencionadas para el diseño de un ecosistema de datos de negocio.

¿Porqué estudiar nuestro diploma?

El Diploma en Data Engineer ofrece una formación integral y práctica para los profesionales que desean dominar las habilidades clave en ingeniería de datos. A través de un enfoque aplicado, los estudiantes aprenderán a diseñar y ejecutar soluciones robustas de procesamiento y gestión de datos en la nube, big data, y visualización, lo que los prepara para enfrentar los desafíos de las organizaciones modernas.



Aprendizaje práctico desde el primer día

Los estudiantes desarrollan proyectos simulando casos reales en cada módulo, lo que les permite aplicar inmediatamente sus conocimientos en entornos de Python, SQL, Big Data, y servicios en la nube.



Domínio de las herramientas más demandadas en el mercado

El diploma incluye formación en tecnologías clave como Apache Airflow, Apache Hadoop, Kafka, y servicios nativos de cloud, preparando a los estudiantes para roles competitivos en la industria de datos.



Proyecto integrador aplicado al negocio

Al finalizar el diploma, los estudiantes diseñan e implementan un ecosistema de datos completo para una empresa o sector de su elección, reforzando su capacidad para resolver problemas empresariales reales.



Visión integral del ecosistema de datos

Los módulos abarcan desde la integración y orquestación de datos hasta la automatización de procesos y la visualización de resultados, ofreciendo una formación completa para manejar el ciclo de vida de los datos.



Certificación

02 Emitida por DMC:

- Especialización en Big Data, por un total de 64 horas académicas.
- Diploma Data Engineer, por un total de 126 horas académicas.

01 Emitida por CertiProf: Big Data Professional Certification (BDPC).



Plana docente de expertos en analítica

Nuestro diploma cuenta con profesionales que se desempeñan en las principales empresas del país y cuentan con una alta experiencia práctica y de liderazgo en data y analytics.



Metodología DMC

Basada en tres pilares "Analiza, Innova y Transforma," combina análisis avanzado, creatividad y aplicación práctica para desarrollar soluciones disruptivas. Preparando a los estudiantes para liderar la transformación digital en el ámbito empresarial.

Malla Curricular

1. Taller de Marca Personal y Empleabilidad

PYTHON FOR DATA ENGINEERING

2. Fundamentos de ETL con Python

- ETL. Definición y herramientas.
- Herramientas Python para ETL.
- Python y sus entornos de ejecución.
- Tipos de datos en Python.
- **Taller:** Creación y ejecución de scripts (programas) en Python (.py)
- API. Definición y librerías para extraer datos.
- **Taller:** Conexión y consulta de datos desde una API.

3. Python para manejo de datos: Pandas. Series y Dataframes

- Pandas. Definición, carga en Python.
- Pandas Series. Características y uso de vectores.
- **Taller:** Operaciones con Series. Búsquedas, Slicing, operaciones aritméticas, tipos de datos.
- Pandas DataFrames. Características y uso de DataFrames. Diferencias respecto a Series.
- **Taller:** Operación con DataFrames. Creación, descripción, visualización.
- **Taller:** Operaciones de agrupación. Agrupaciones directas y por Agregación simple y múltiple (varios campos).
- **Taller:** Guardar DataFrames en archivos planos (Json y CSV) y base de datos (MySQL).

4. Conexión a bases de datos relacionales

- **Taller:** Carga de datos desde un archivo plano, procesamiento y descarga en una base de datos.

SQL FOR DATA ENGINEERING

5. El Lenguaje Transact-SQL

- SQL y T-SQL. Definición, diferencias.
- Lenguaje de definición de datos (DDL). Definición, alcance y comandos asociados (CREATE, ALTER, DROP).
- Lenguaje de manipulación de datos (DML). Definición, alcance y estructura del comando SELECT...FROM...
- **Taller:** Estructura básica de una consulta. Uso de SELECT...FROM...
- **Taller:** Consultas condicionales, uso de WHERE y operadores lógicos.
- **Taller:** Consultas de agregación, uso de GROUP BY, COUNT, MAX, MIN, SUM, AVG,
- **Taller:** Pivoteo de tablas, uso de PIVOT.
- **Taller:** Consultas multi-tabla. Uso del comando JOIN y variantes (LEFT, RIGHT, FULL)
- **Taller:** Operadores de conjunto, uso de UNION, INTERSECT, EXCEPT.

Malla Curricular

6. Transact-SQL Avanzado

- **Taller:** Transformación y operación de columnas, uso de operadores aritméticos, funciones de fechas, funciones de textos, uso de IIF, ISNULL, NULLIF.
- **Taller:** Filtrado avanzado, uso de IN, ANY AND SOME, ALL, EXISTS.

- **Taller:** Conversión de tipos de datos, uso de CAST, CONVERT, FORMAT, PARSE.
- **Taller:** Encapsulamiento de consultas en Procedimientos almacenados. Uso de estructuras condicionales y bucles.

7. Tópicos Avanzados sobre Tablas

- **Taller:** Uso de tablas derivadas y tablas temporales para pre-procesamiento de datos.
- Instrucciones tipo Query jerárquica y correlacionales.
- **Taller:** Manejo de expresiones de tablas (CROSS APPLY, OUTER APPLY).

- **Taller:** Uso de cursores para copia y transferencia de datos.
- **Taller:** Automatización de scripts con Jobs (schedule).
- **Taller:** Diseño de un script ETL con lenguaje T-SQL (De BD a BD)

BIG DATA PROCESSING

8. 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.

9. 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.

10. Procesamiento distribuido con Apache Spark

- Introducción a Apache Spark. Definición, casos de uso.
- Funciones con PySpark.
- Dataframes y RDDs.
- Funciones de Apache Spark.
- **Taller:** Extracción y transformación de datos en Apache Spark.

11. Ingeniería de datos con Databricks

- Introducción a Databricks. Versión independiente vs. On-cloud. Casos de uso.
- **Taller:** Reconocimiento de la interfaz de Databricks.
- Data lakehouse vs Datalake.
- Iceberg vs Delta lake vs Apache Hudi.
- **Taller:** Creación de un Lakehouse con arquitectura Medallion con Delta Lake.

Malla Curricular

12. Real-time Fundamentals

- Introducción a Apache Kafka y Kafka Connect. Definición, casos de uso.
- Arquitectura Publicador – Suscriptor.
- Conectores vs Debezium.
- **Taller:** Configuración de un conector para escuchar una base de datos local.

- **Taller:** Configuración de un conector para escuchar una base de datos en nube.
- **Taller:** Configuración de un conector para escuchar un archivo plano y una carpeta.
- **Taller:** Configuración de Apache Kafka para captura de datos que se almacenan en una base de datos relacional local.

CLOUD DATA ENGINEERING

13. Fundamentos de Cloud Computing

- Definición, modelo de costos, ventajas de su uso.
- Principales proveedores. Regiones y zonas de disponibilidad.
- Repaso de conceptos generales: Virtualización, Uso bajo demanda, despliegue escalable.
- Tipos de servicios. IaaS, PaaS y SaaS.

14. Introducción a la Ingeniería de datos con AWS

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

15. Introducción a la Ingeniería de datos con Azure

- **Taller:** Exploración de la consola de Azure.
- Exploración y descripción de los 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.

16. 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.

Malla Curricular

DATA INTEGRATION & ORCHESTRATION

17. Apache Airflow

- Apache Airflow. Definición.
- DAG (Direct-Acyclic-Graph). Definición y casos de uso.
- Uso de Scheduler.
- Task and Operator.

• **Taller:** Implementación de un DAG con Bash Operator.

- **Taller:** Implementación de un DAG con Python Operator.
- **Taller:** Implementación de un DAG con Apache Spark Operator.

WEB SCRAPING CON PYTHON

18. Web Scraping Fundamentals

- Web Scraping. Definición, debate legal-ético.
- Métodos de extracción de datos.
- Técnicas de web scraping. Estático, Requests, Dinámico.

- **Taller:** Scraping a un sitio web y almacenamiento de su contenido en archivo plano.

DATAOPS

19. Fundamentos de DataOps

- DataOps. Definición y características.
- Devops vs DataOps.
- Conceptos asociados: Continuos Delivery (CD) y Continuos Integration (CI)
- Servicios de automatización de despliegue: Jenkins, Azure DevOps, Github Actions, Gitlab CI).

- Fundamentos de Infraestructura como código. Definición y características de Terraform.

20. Git and GitHub

- El control de versiones. Definición y características.
- Git. Definición, principales comandos.
- GitHub. Definición.
- Configuración de llaves SSH.
- Directorio de trabajo.

- **Taller:** Uso del comando Commit.
- **Taller:** Fusionar y el comando Merge.
- **Taller:** Ramas y el comando Branch.
- **Taller:** Creación de un repositorio en GitHub.

Malla Curricular

DATA VISUALIZATION

21. Data Visualización Fundamentals

- Inteligencia de negocios y herramientas de visualización.

- **Taller:** Exploración de Power BI y conectores disponibles.

- **Taller:** Conexión de Power BI a fuentes de datos locales y cloud (Azure).

- **Taller:** Transformación básica de datos con Power Query.

- **Taller:** Implementación de un Dashboard básico empleando fuentes locales y cloud.

PROYECTO INTEGRADOR

- **Sesión 1.** Lineamientos generales.
- **Sesiones 2-3.** Revisión de avance y feedback.

- **Sesión 4.** Presentación final y sustentación.

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.



Proyecto Integrador

Combina las herramientas y conocimientos adquiridos en un proyecto aplicado a casos reales.

Docentes Expertos

Aprende con los líderes de las mejores empresas de Latam.



Oscar Barrientos

Chief Operations
Officer



Angel Tintaya

Senior Data
Engineer



Juan Salinas

Senior Data
Engineer



Miguel García

Arquitecto de
Seguridad de TI



Jenifer Garate

Sr Integration
Engineer



Tony Trujillo

Data
Architect



*En caso de contingencias podría cambiar alguno de los docentes por otro profesional de similar perfil.



Certificación Internacional

Al completar el Diploma, también contarás con la oportunidad de rendir el **Big Data Professional Certificate BDPC** emitido por Certiprot, por solo **20 dólares (opcional)**.

El costo real del examen Certiprot es de 150 dólares

El alumno tendrá hasta 02 oportunidades de aprobar el examen por el pago realizado.



* Para acceder al beneficio el alumno debe aprobar satisfactoriamente la capacitación.

* El alumno tiene libre disposición a decidir de aplicar o no a la certificación.

¿Por qué elegirnos?

Docentes líderes



Nuestros docentes son destacados expertos en data & analytics que lideran equipos de alto rendimiento en las empresas más grandes de Latam.

Portafolio especializado



Tenemos el portafolio más completo con +150 capacitaciones sincrónicas y asincrónicas que se ajustan a diferentes perfiles y niveles de conocimiento.

Metodología innovadora



Nuestra metodología "Aprende haciendo" ha logrado que nuestra comunidad de +25K profesionales en todo Latam mejoren su situación laboral.

Reconocimiento empresarial



Las empresas worldclass de Latam confían en nosotros para acompañarlas en su transformación hacia el enfoque data driven desde hace 15 años.

Atención personalizada



Desde que te matriculas recibirás atención y asesorías en todo lo que necesites para que aproveches al máximo tu inversión.





www.dmc.pe