

Diploma

DATA ENGINEER



Presentación

Los recientes avances en materia de Ciencia de datos e inteligencia artificial en general han promovido el crecimiento exponencial de los datos almacenados por las empresas. Ahora la data no solo proviene desde los sistemas transaccionales, sino desde la web, desde sensores, o, inclusive, del tratamiento que se le da a la data ya existente.

El Diploma Data Engineer te permitirá diseñar e implementar todo un ecosistema de datos extrapolable a la gran mayoría de negocios y sectores empresariales, y que harás posible a través del uso y aplicación de diversos lenguajes de programación, gestores de bases de datos, Big Data y otras tecnologías emergentes.



Sobre este Diploma

43

sesiones

168

horas
académicas

47

talleres
prácticos

01

proyecto para
tu portafolio

¿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 y blandas.**
- **Mentoría especializada** con docentes expertos.
- Acompañamiento **constante.**



¿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.

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.
- Repaso de conceptos generales: Virtualización, Uso bajo demanda, despliegue escalable.
- Principales proveedores. Regiones y zonas de disponibilidad.
- 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.

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



Docentes Expertos

Aprende con los líderes de las mejores empresas de Latam.



Juan Salinas

Data Solutions Expert |  NOVACOMP

Profesional en Ingeniería de Sistemas, con más de 9 años de experiencia en áreas de tecnología de la información. Actualmente lidero el equipo de Datos y Analítica de Seguros en Scotiabank, y también desempeño funciones de consultoría y docencia en Computación en la Nube, Big Data y Analítica.

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

Certificación DMC INSTITUTE

Por aprobación del Diploma Data Engineer, por un total de 168 horas académicas.



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.



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:

