Diploma

ADVANCED DATA ENGINEER

DMC | INSTITUTE

Presentación

El Diploma Advanced Data Engineer te preparará y potenciará tus habilidades en ingeniería de datos, permitiéndote dominar las técnicas para la mejora y optimización de sistemas dedicados que soporten Data Warehouse y Big Data, así como diseñar e implementar soluciones en la nube bajo el enfoque Data Fabric, que permite crear una única fuente de verdad organizacional, integrando y consolidando todos los silos de información.

Sobre este Diploma

37
sesiones

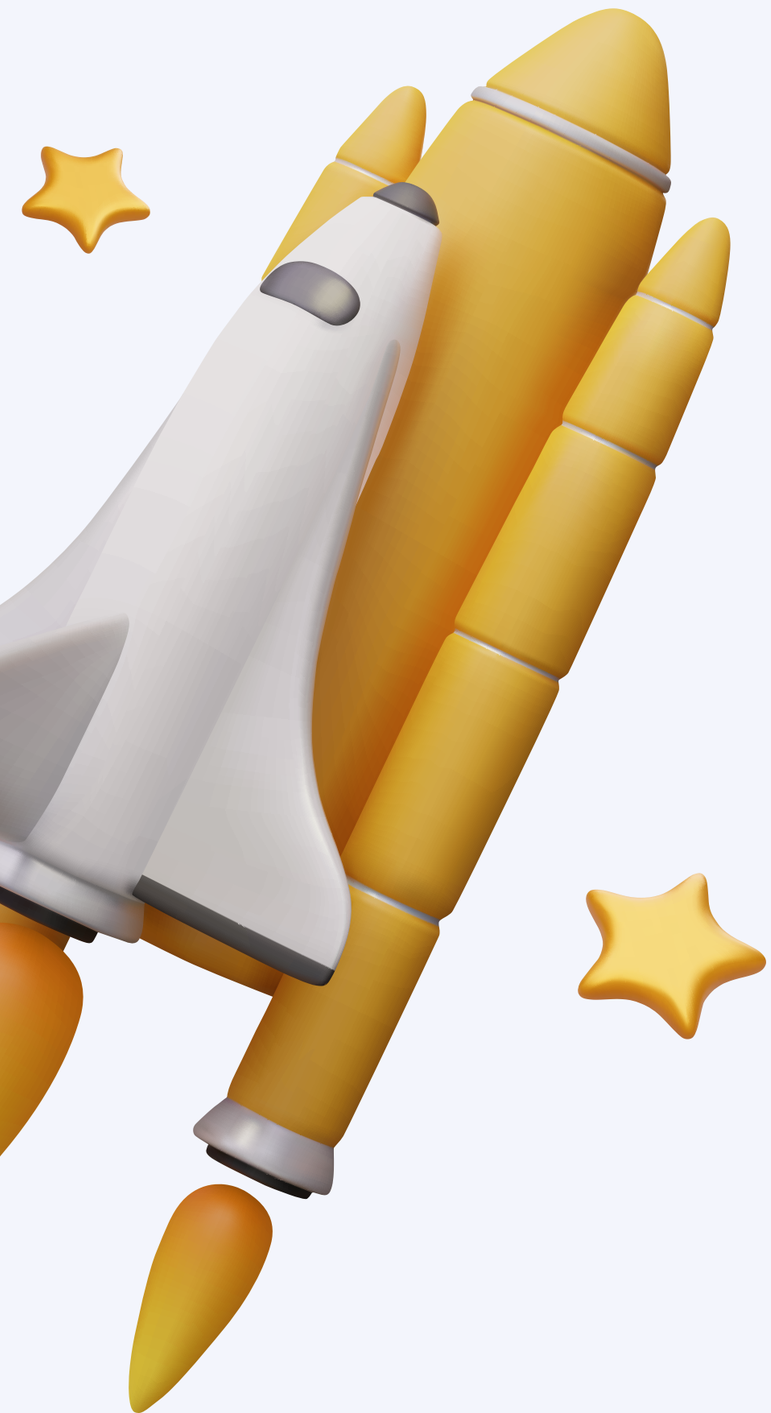
144
horas académicas

47
talleres prácticos

01
proyecto para tu portafolio

¿Cómo impulsamos tu carrera?

- Sesiones 80% **enfocadas en la práctica.**
- Enfoque en **Casos Reales** enfrentando los retos del mercado.
- Énfasis en **habilidades técnicas y blandas.**
- **Mentoría especializada** con docentes praticioners.
- Acompañamiento **constante.**



¿Porqué estudiar este diploma?

Lleva tu potencial en ingeniería de datos a su máximo nivel y lidera el desarrollo de soluciones robustas que integren todos los datos de la organización y basadas en el enfoque Data Fabric.

Culmina tu capacitación con una solución avanzada de ingeniería de datos construida por ti mismo.

Aprende a mejorar y a optimizar sistemas dedicados (on-premise) que operen sobre SQL Server, así como aquellos basados en Apache Spark (Big Data).

Aprende a diseñar a implementar soluciones Data Fabric sobre las plataformas cloud líderes del mercado: Azure, AWS y GCP.

¿En qué se diferencia esta versión respecto al tradicional Diploma Data Engineer?

- Este Diploma se orienta al scraping de datos avanzado mediante el enfoque y características de la librería Selenium, que ofrece mayores prestaciones y parámetros de configuración para proyectos más robustos.

- Este Diploma se orienta a la optimización y performance de repositorio de datos en entornos dedicados, como son Data Warehouse y Data Lakes, en vez de su diseño o implementación básica.

- Este Diploma emplea las plataformas cloud para la implementación de arquitecturas de datos basadas en el enfoque Data Fabric.

- Este Diploma se orienta a la orquestación de entornos dedicados y entornos cloud mediante la plataforma de Apache Airflow.

Objetivo del diploma

- Comprende el impacto que tiene el desarrollo de su marca personal para su vida profesional.
- Implementa Data Warehouses de tres capas automatizados con paquetes SSIS.
- Utiliza Python para la implementación de componentes para soluciones de ingeniería de datos.
- Aplica técnicas en el entorno Apache Spark para garantizar la velocidad de respuesta y la optimización de recursos de cómputo en la ejecución de ETL que operan sobre un entorno Big Data.
- Implementa soluciones de ingeniería de datos en el entorno de Ms. Azure, AWS Cloud y GCP.
- Utiliza Apache AirFlow para la implementación y calendarización de tareas

Objetivo Final

El alumno desarrolla una solución de ingeniería de datos con dos características generales a) híbrida, es decir que integra entornos dedicados (on-premise) y cloud, y b) basada en Data Fabric, la misma que el alumno construye en un entorno Azure, AWS o GCP.

¿A quién está dirigido?

1. Ingenieros de datos junior y semi-senior

Profesionales que actualmente desempeñen este rol, y busquen:
- Optimizar sus entornos de datos dedicados, tanto Data Warehouses sobre SQL Server como Big Data sobre Apache Spark.
- Llevar sus pipelines de datos construidos en Cloud a un enfoque Data Fabric y orquestarlos con sus sistemas locales.

2. Analistas y científicos de datos con perfiles técnicos relacionados a las carreras de informática, sistemas o afines

Personas que desempeñen alguno de estos roles, y busquen:
- Migrar a puestos abocados al diseño e implementación de pipelines de datos y ETLs.

¿Cuáles son los requisitos?

Conocimientos / Habilidades	Experiencia Laboral	Tecnológicos
• Conocimiento de programación en lenguaje Python (nivel avanzado) • Conocimiento de operación y gestión de entornos Big Data. • Conocimiento de plataforma cloud y gestión de servicios.	• Mínimo un año de experiencia laboral como ingeniero de datos o haciendo sus veces.	• Contar con una laptop o computadora de escritorio con disponibilidad de micrófono y cámara web. • Tener instalado los softwares y herramientas señalados en la sección Contenidos.

Perfil del egresado

El egresado del Diploma Advanced Data Engineer estará en la capacidad de implementar soluciones de ingeniería de datos que integran sistemas dedicados (on-premise) y en nube sobre alguna de las plataformas líderes del mercado como Azure, AWS o GCP, considerando técnicas avanzadas para la ingesta de datos y control de performance.

Campo Laboral

El egresado del Diploma Advanced Data Engineer podrá laboral en puestos relacionados a:

- Ingeniero de datos / Data Engineer
- Ingeniero de datos cloud / Cloud Data Engineer.
- Big Data Engineer.

Herramientas

Google Colab

Apache Spark

Azure

AWS

SQL Server

Google Cloud Platform

Apache Airflow

Malla Curricular

I. Taller de Marca y Empleabilidad

- Actividad de sociabilización y contacto.
- Marca personal ¿Qué es y cómo desarrollarla?
- Empleabilidad y ser empleable. Diferencias clave.
- ¿Cómo hacer más atractivo el curriculum?
- ¿Cómo afrontar una entrevista de trabajo?

II. Diseño de Data Warehouses avanzados con SQLSERVER y SSIS

1. Arquitectura y modelamiento de componentes de ingeniería de datos

- Modelamiento dimensional: Star & Snowflake schemas.
- Arquitectura moderna de Data Warehouses.
- Arquitectura moderna de Data Lakes.
- Arquitectura moderna de Data Lakehouses.

2. Implementación de Data Warehouses avanzados (SQLSERVER + SISS)

- Introducción a SQL Server Integration Services (SSIS).
- **Taller:** Diseño de ETL básicos con SSIS.
- **Taller:** Implementación de DWH con SQL Server: Capa Stage.
- **Taller:** Implementación de DWH con SQL Server: Capa Operational Data Store (ODS).
- **Taller:** Implementación de DWH con SQL Server: Capa Business Data Store (BDS).
- **Taller:** Automatización de DWH con paquetes SSIS y jobs.

3. Configuración de Data Warehouses avanzados

- Técnicas de Ingesta de datos: Change Data Capture (CDC), Delta ingestion, Full ingestion.
- **Taller:** Implementación de las técnicas de ingesta de datos con SQL Server.
- Técnicas de optimización del rendimiento del almacenamiento: Indexing, partitioning, bucketing, design by query, clustering).
- **Taller:** Implementación de las técnicas de optimización con SQL Server.

III. Ingesta y procesamiento avanzado con Python

4. Control de bases de datos

- Repaso de principales librerías para procesamiento de datos.
- ORM SQLAlchemy. Definición, usos en ingeniería de datos.
- **Taller:** Diseño de un ETL sencillo con pandas y SQLAlchemy.

5. Control de concurrencia de datos

- IO-Bound vs. CPU-Bound. Aplicaciones en ingeniería de datos.
- El método ThreadPoolExecutor de Python. Casos de uso, secuencia de implementación.
- **Taller:** Uso de ThreadPoolExecutor para procesamiento paralelo de archivos.
- **Taller:** Uso de ThreadPoolExecutor para carga de datos desde un API.
- **Taller:** Uso de ThreadPoolExecutor para ejecución en paralelo de consultas a bases de datos.
- **Taller:** Solución E2E, de API a base de datos relacional.

6. Scraping avanzado con Selenium

- Fundamentos Web Scraping.
- La librería Selenium. Casos de uso, características.
- Cómo evitar las trampas en el proceso del Web Scraping (Ajustando encabezados, Manejo de cookies con JavaScript, Huellas dactilares TLS, Timing Is Everything).
- Expresiones regulares.
- Web Scraping in Parallel.
- Web Scraping Proxies.
- **Taller:** Obtener datos de un archivo PDF utilizando expresiones regulares.
- **Taller:** Web Scraping a sitios web.

IV. Optimización de procesos Big Data

7. Repaso Apache Spark en Ingeniería de datos

- Big Data en proyectos de ingeniería de datos.
- Apache Spark para Big Data.
- Apache Spark. Módulo, Dataframe API y PySpark Functions.
- **Taller:** Implementación de un ETL básico con PySpark.

8. Apache Spark for Tunning and Performance

Análisis de consultas Spark

- El Spark Query Plan. Definición, uso en la etapa de optimización de Spark.
- **Taller:** Creación e interpretación de Spark Query Plan.
- **Taller:** Creación e interpretación de Spark DAGs.

Ajuste de recursos

- **Taller:** Monitoreo y ajuste del consumo de memoria con Spark UI (Memory management)
- **Taller:** Ajuste de executors para la optimización de memoria y núcleos (Executor tuning).
- **Taller:** Ajuste de particiones para la optimización de disco y red.

Malla Curricular

Optimización avanzada
• **Taller:** Mejora de rendimiento basada en partitioning (particionamiento lógico).
• **Taller:** Mejora de rendimiento basada en Buckets.
• **Taller:** Mejora basada en Caching (memoria caché).
• **Taller:** Mejora basada en balance de carga de nodos (Solve Data Skew). Salting, Broadcast joins.
• **Taller:** Mejora basada en reducción dinámica de particiones de datos (Dynamic Partition Pruning).

V. Azure para ingeniería de datos

9. Fundamentos de ingeniería de datos con Azure
· Principales Servicios de Azure para Ingeniería de datos. Denominación y modelos de costo asociados.

10. Arquitectura "Data Fabric" en Azure
· Visión Integral de la Arquitectura Data Fabric.
· Enfoques arquitectónicos Data Fabric y Data Mesh.
· Servicios de Azure asociados a Data Fabric.
Discovering & Ingest
· Planteamiento de la casuística a resolver.
• **Taller:** Revisión de orígenes de datos y elaboración del catálogo de datos con Azure Data Catalog.
• **Taller:** Configuración de Azure Data Factory para ingestar datos de diversas fuentes.
Data Lakehousing
· Introducción a los servicios Azure Blob Storage y Storage Account.
• **Taller:** Implementación de un Data Lake con Azure Data Lake Storage. Integración con Data Factory.
Big Data Pipeline
· Introducción a Databricks y Azure Databricks.
• **Taller:** Uso de Azure Databricks para crear un pipeline para procesar y transformar lo almacenado en el Data Lake.
Data Warehousing
· Introducción a Azure Synapse Analytics.
• **Taller:** Uso de Azure Synapse Analytics para implementar un Data Warehouse. Integración con Azure Databricks para consumo de datos del Data Lake.
Data Deliver
· Introducción a Azure Analysis Services.
• **Taller:** Uso de Azure Analysis Services para implementar un Datamart. Integración con Azure Synapse Analytics.
• **Taller:** Conexión con Power BI y test de funcionamiento.

VI. AWS para ingeniería de datos

11. Fundamentos de ingeniería de datos con AWS
· Principales Servicios de AWS para Ingeniería de datos. Denominación y modelos de costo asociados.

12. Arquitectura Data Fabric en AWS
· Servicios de AWS asociados a Data Fabric.
Discovering & Ingest
· Planteamiento de la casuística a resolver.
• **Taller:** Revisión de orígenes de datos y elaboración del catálogo de datos con AWS Glue Data Catalog.
• **Taller:** Configuración de AWS Glue DataBrw y AWS Glue para ingestar datos de diversas fuentes.
Data Lakehousing
· Introducción al servicio S3 y seguridad en Buckets.
• **Taller:** Implementación de un Data Lake mediante instancias S3. Integración con Glue.
Big Data Pipeline
· Introducción al servicio AWS EMR.
• **Taller:** Uso de AWS EMR para crear un pipeline para procesar y transformar lo almacenado en el Data Lake.
Data Warehousing
· Introducción a Amazon Redshift.
• **Taller:** Uso de Amazon Redshift para implementar un Data Warehouse. Integración con EMR.
Data Deliver
• **Taller:** Revisión general de Amazon QuickSight.
• **Taller:** Test de conexión desde Power BI.

VII. GCP para ingeniería de datos

13. Fundamentos de ingeniería de datos con GCP
· Principales Servicios de GCP para Ingeniería de datos.

14. Arquitectura Data Fabric en GCP
Discovering & Ingest
· Planteamiento de la casuística a resolver.
· Introducción a los servicios para ingesta de datos: Cloud Data Fusion, Cloud Dataflow y Pub/Sub.
• **Taller:** Configuración de Cloud Data Fusion para ingestar datos de diversas fuentes.
Data Lakehousing
· El servicio Cloud Storage y la gestión de su seguridad.
• **Taller:** Implementación de un Data lake mediante Cloud Storage.

Malla Curricular

Big Data Pipeline
• Introducción a Dataproc.
• **Taller:** Uso de Dataproc para procesar y transformar lo almacenado en el Data lake.
Data Warehousing
• Introducción a BigQuery.
• **Taller:** Uso de BigQuery para implementar un Data Warehouse. Integración con Dataproc.
Data Deliver
• **Taller:** Revisión general de BigQuery BI Engine.
• **Taller:** Test de conexión desde Power BI.

VIII. Entornos híbridos: Dedicados y Cloud

15. Introducción a Apache AirFlow
• Apache Airflow. Definición, casos de uso en ingeniería de datos.
• Estructura de un notebook de Airflow.
• **Taller:** Implementación y calendarización de una tarea que carga de datos desde una API de clima.

16. Apache AirFlow para ingeniería de datos
• **Taller:** Implementación y calendarización de una tarea para carga y transformación de datos de SQL Server local a servicio de almacenamiento Azure.

IX. Proyecto integrador

1. Componente: Data Warehouse de la solución - Módulo II
• Implementado con SQL Server.
• Una técnica de ingesta de datos.
• Aplicación de dos técnicas de optimización (mínimo).

2. Componente: Ingesta de datos de la solución - Módulo III
• Implementado con código Python.
• Ingesta basada en API.
• Ingesta basada en Scraping.
• Por lo menos uno de ellos con ThreadPoolExecutor.

3. Componente: ETLs optimizados sobre Apache Spark (Big Data) - Módulo IV
• Implementados con código Python y PySpark.
• Aplicación de dos técnicas de optimización (mínimo).

4. Componente: Data Fabric en Ms. Azure - Módulo V
• Flujo completo ajustado a los recursos y objetivos del proyecto.

5. Componente: Data Fabric en AWS - Módulo VI
• Flujo completo ajustado a los recursos y objetivos del proyecto.

6. Componente: Data Fabric en GCP - Módulo VII
• Flujo completo ajustado a los recursos y objetivos del proyecto.

3. Componente: AirFlow Scheduled Pipeline - Módulo VIII
• Implementado con Apache Airflow.
• Debe realizar cargas programadas del entorno dedicado (on-premise) a Data Fabric.

DMC | INSTITUTE

Certificación DMC

Por aprobación del Diploma Advanced Data Engineer, por un total de 144 horas académicas.

DMC | INSTITUTE

DIPLOMA

ESPECIALISTA EN ADVANCED DATA ENGINEER

Se otorga a

NOMBRES Y APELLIDOS

Luego de completar el diploma que se realizó del 00 de mes al 00 de mes, completando un total de **144 horas académicas.**

Mes del 2025

MBA Jonny Chambi
GERENTE GENERAL

MBA Dionicio Velasquez T.
COORDINADOR ACADÉMICO

00000

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

DMC | INSTITUTE

Certificación Internacional

Al finalizar la capacitación, tendrás la oportunidad de acceder al

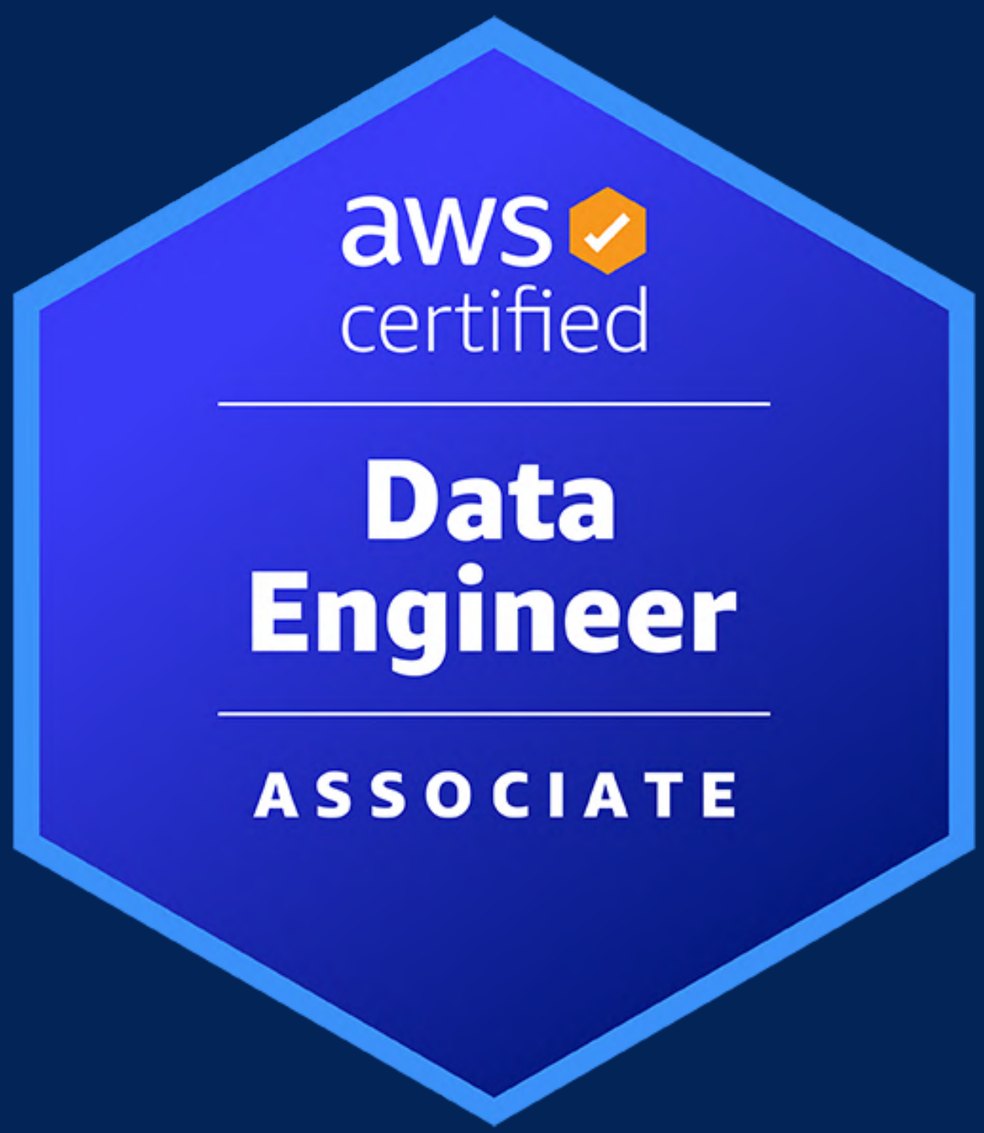
AWS CERTIFIED DATA ENGINEER - ASSOCIATE

con un descuento especial

Costo del examen
150 dólares

50%

Precio con descuento
75 dólares

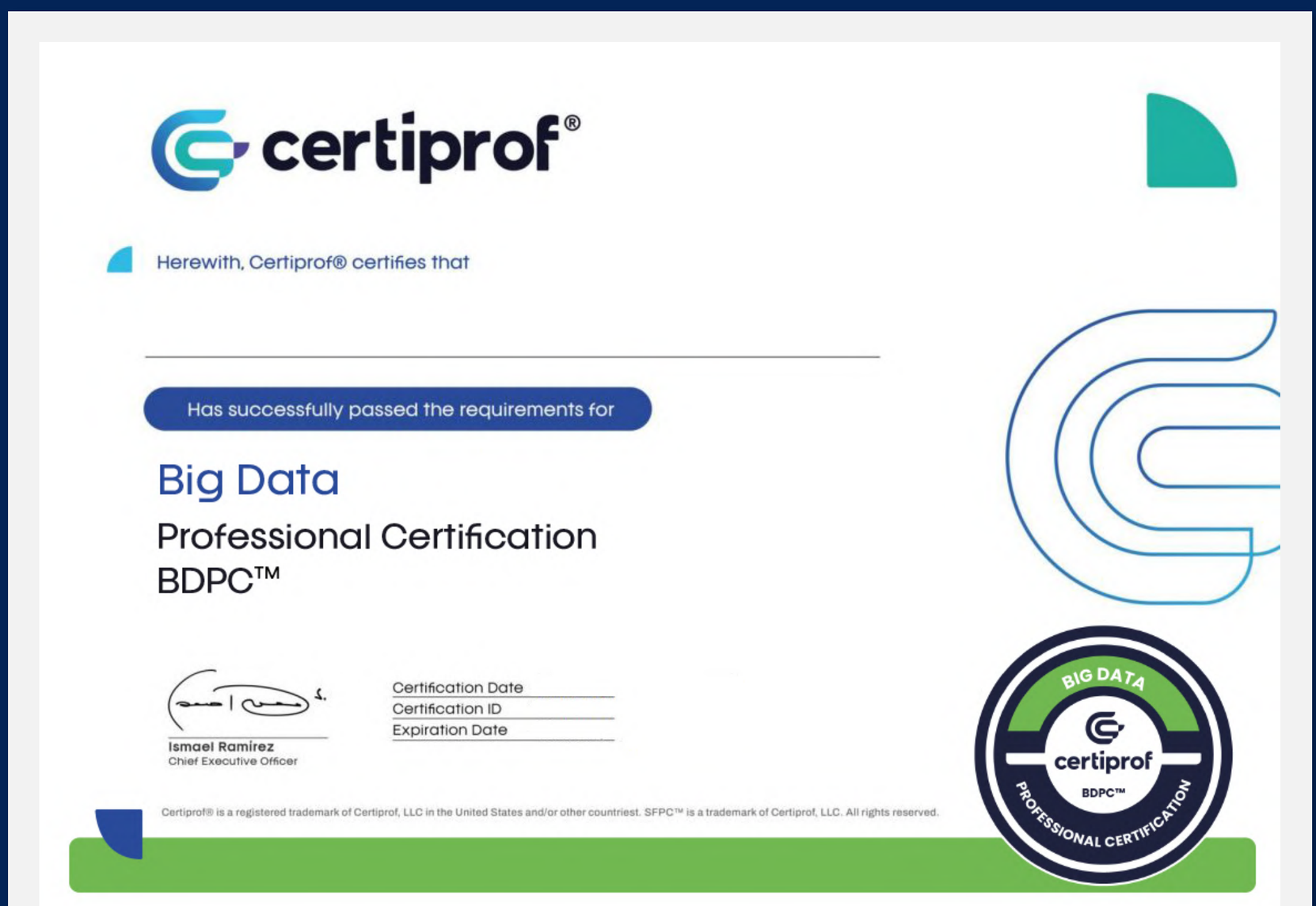


DMC | INSTITUTE

Certificación Internacional

Al finalizar la capacitación, tendrás la oportunidad de acceder
al **Big Data Professional Certification (BDPC)**
de CertiProf con un descuento especial.

Costo del examen		Precio con descuento
150 dólares	**67.33%**	**49 dólares**



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!

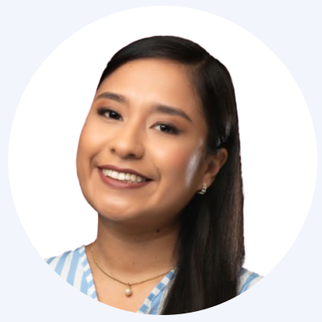
• Tú eliges si aplicar o no
El beneficio es opcional. Solo tú puedes decidir aprovecharlo

•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.

Docentes Expertos

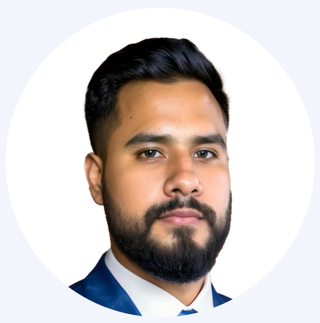
Miguel Balcazar

Senior Lead, Data Engineering
en KYNDRYL



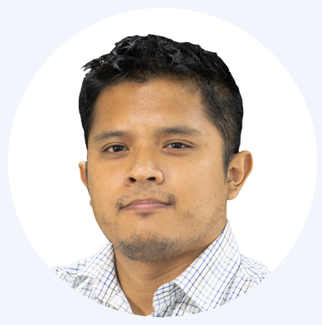
Maria Moran

Employability Analyst
en PRONABEC



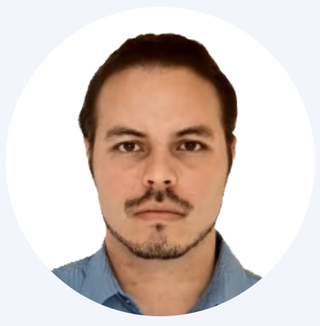
Antony Alza

Analista Sr. de Ingenieria Financiera
en ALICORP



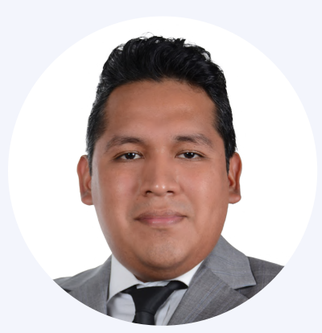
Tony Trujillo

Arquitecto de datos
en IDATHA



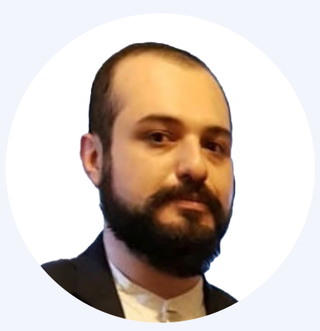
Miguel García

Especialista en TI
en E2E SOLUTIONS



Angel Tintaya

Senior Data Engineer
en TRANZACT



Gustavo Rangel

Google Cloud Platform Specialist
en GOOGLE

Importante:

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

DMC | INSTITUTE

¿Por qué elegirnos?

+16 Más de 16 años de experiencia.

+300 Más de 300 empresas asesoradas en Perú, Ecuador y Bolivia.

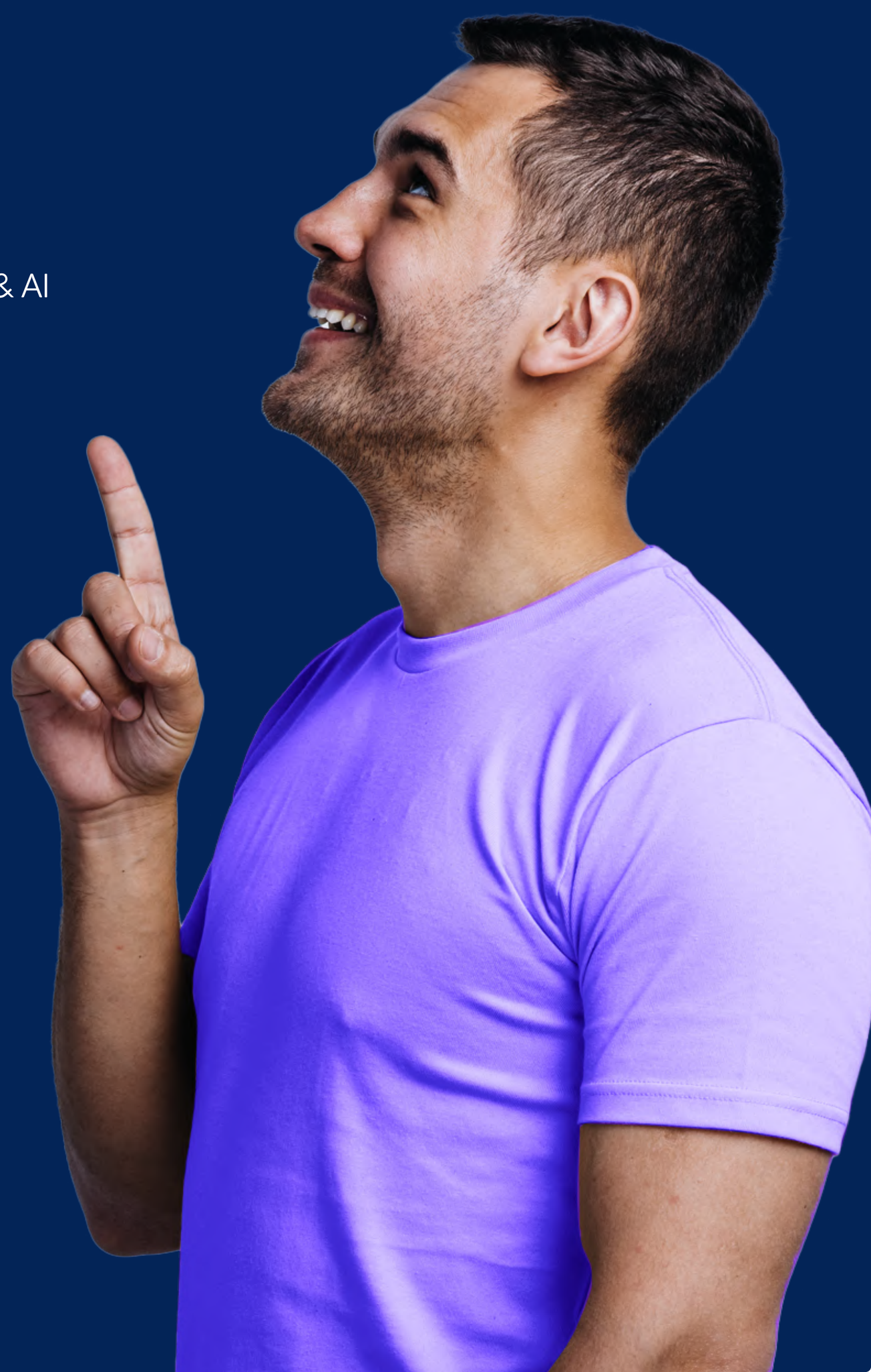
35k 35 mil profesionales capacitados en más de 20 países de América Latina.

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 exlusivos: Networking, empleabilidad, habilidades blandas.

Excelente nivel de servicio.



Nuestros Partners

CertiProf® | Partner

Google Partners

aws partner network

PYTHON INSTITUTE
Open Education & Development Group

Microsoft Partner

Estas empresas confían en nosotros

BBVA

ANTAMINA

PROM PERÚ

SCOTIABANK

PACÍFICO SEGUROS

SUNAT

CAJA HUANCAYO

BUENAVENTURA

PRONABEC

CAJA AREQUIPA

RIMAC

BCRP

MIBANCO

MAPFRE

ONCOSALUD

LOS ANDES

Métodos de pago

J&J DATA MINING CONSULTING S.A.C.
RUC: 20520972740

1. Depósito en cuenta BCP

- Corriente soles BCP: **193-225-1181-0-01**
- CCI BCP: **00219300225118100116**
- Corriente BCP dólares: **193-2318515-1-52**
- CCI BCP dólares: **002-193-002318515152-11**

2. Depósito en cuenta BBVA

- Ahorros BBVA soles: **0011-0177-02-00180473**
- CCI BBVA: **011-177-000200180473-37**

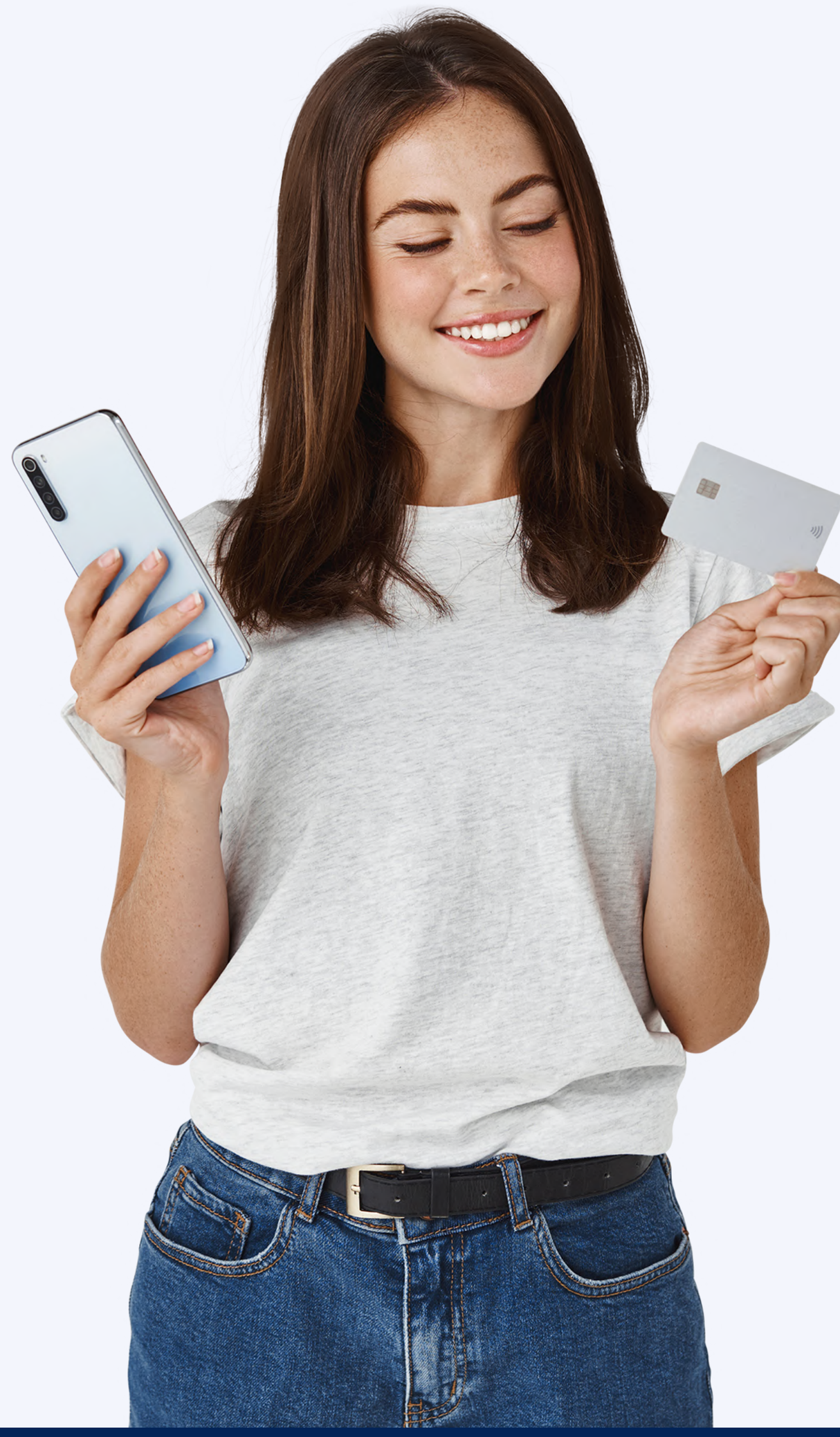
2. Pago Online

Generamos un link de pago online donde se acepta todas las tarjetas.

3. Pago con Yape

A nombre de J J Data Mining Consulting Sac

4. Pago online por PayPal



06 CUOTAS SIN INTERESES pagando con:

DMC | INSTITUTE

Analiza, aplica, ***transforma***

Visita nuestra web

www.dmc.pe

Síguenos en: