

Máster Internacional en

# DATA ENGINEERING ON CLOUD



Inicio  
27 DE NOVIEMBRE, 2024

Duración  
11 MESES

Inversión  
S/. 15,990

📡 Online  
en Vivo

**Horario:** Lunes y miércoles de 7:30 pm a 10:30 pm

# CONOCE EL MASTER INTERNACIONAL EN DATA ENGINEERING ON CLOUD

El uso y explotación de los datos es transversal a las áreas o puestos de trabajo de una organización y la democratización de estos se ha vuelto crítica para la toma de decisiones, el diseño de estrategias y la creación de ventaja competitiva. Frente a este panorama, la demanda de profesionales que estén en la capacidad de poner los datos a disposición de todo aquel que lo necesite en la compañía y, sobre todo, a partir de soluciones basadas en las mejores alternativas tecnológicas, se ha vuelto indispensable.

Por ello DMC Perú, presenta el Master en Data Engineering on Cloud, que te capacitará en el planeamiento, diseño e implementación de soluciones de ingeniería de datos totalmente alineadas a los grandes objetivos corporativos, ya sea que estos requieran de proyectos de analítica o de ciencia de datos, con datos estructurados o no estructurados, on-premise u on-cloud.



# ¿POR QUÉ ESTUDIAR NUESTRO MASTER?



## Plana docente internacional

Contamos con docentes que trabajan en empresas líderes a nivel Latam y que podrán compartir sus conocimientos desde una visión global además de profundizar en las nuevas tendencias del mercado.



## Ofrecemos triple certificación

### 02 emitidas por DMC:

- Máster Internacional en Data Engineering on Cloud
- Especialización en Cloud Data Engineering

### 01 emitida por Microsoft Azure:

- Microsoft Certified: Azure Data Fundamentals



## Data Engineer Challenge

El alumno podrá desarrollar 02 proyectos donde creará un Data Warehouse y un Data Lakehouse on Cloud que lo ayudará a fortalecer su portafolio.



## Mentoring 1 to 1

El alumno recibirá sesiones personalizadas para su desarrollo profesional por parte de un profesional destacado de la industria.



## 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.



## Jornada Internacional en Data & Analytics

Evento diseñado para reunir a profesionales de distintos países a compartir sus conocimientos en el campo de data science y analítica.

# OBJETIVO PRINCIPAL

El máster tiene como objetivo principal:

Desarrollarás proyectos de ingeniería de datos con enfoque en negocios, considerando tanto soluciones para datos relacionales mediante Data Warehouses robustos, así como aquellas para datos mixtos mediante Data Lakes y servicios nube que garanticen la democratización de los datos y la continuidad del negocio.

# PERFIL DEL EGRESADO

El egresado de la máster será capaz de:

1. Planificar, diseñar e implementar Data Warehouses que den soporte a los proyectos de datos actuales y futuros de la organización (Data Roadmap), empleando herramientas y técnicas de vanguardia de la arquitectura, modelamiento, gobierno y seguridad de datos.
2. Planificar, diseñar e implementar Data Lakehouses en entornos Cloud que den soporte a los proyectos de datos actuales y futuros de la organización (Data Roadmap) y con sustento financiero, empleando herramientas tanto propietarias como open-source de vanguardia en la industria.

# ¿A QUIÉN ESTÁ DIRIGIDO?

- **Ingeniero de datos**

Personas que actualmente se desempeñen en el campo de la ingeniería de datos con más dos años de experiencia, y que busquen:

- Incorporar a su expertirse técnica, habilidades de negocio que le permitan alinear sus soluciones a los objetivos organizacionales.
- Puestos de mayor jerarquía que les permitan gestionar proyectos de ingeniería de datos.

- **Supervisores de bases de datos y DBAs**

Personas que desempeñan estos roles y buscan:

- Migrar sus centros de almacenamiento y gestión de datos a entornos cloud, a través de la aplicación de las mejores prácticas de arquitectura, gobierno, seguridad e ingeniería de datos.



## REQUISITOS PARA POSTULAR

- Sólidos conocimientos en programación Python y manejo de entornos Windows y Linux.
- Experiencia laboral mayor a dos años desarrollando proyectos de análisis o ingeniería de datos.

# MALLA CURRICULAR



# CONTENIDO

## Track 01: Corporative Data Warehousing

### Módulo I: Data Team Leadership

- Me conozco. Autoconocimiento y descubrimiento del líder que hay en mí.
  - Colaboro. Engagement de equipos.
  - Lidero. Escucho, inspiro y guío.
- 

### Módulo II: Data Strategy Plan

#### Estrategia de Datos

- Definición, enfoque e impacto en las organizaciones.
  - Roles en equipos de data y analytics.
  - Artefactos para el diseño de una estrategia de datos.
  - Recomendaciones y buenas prácticas en el diseño de una estrategia de datos.
  - **Taller: Diseño del roadmap de una estrategia de datos.**
- 

### Módulo III: Data Architecture & Modeling Fundamentals

#### Data Engineering Lifecycle

- Evolución de la Ingeniería de datos.
- ETL vs ELT.
- Arquitectura de una solución BI.
- Modelos Multidimensionales.
- Tecnologías Data Storage y Data Process.
- **Taller: Desarrollo de un diagrama de una arquitectura BI.**

#### Modern Data Architecture

- Arquitectura de una solución Big Data.
- Taxonomía del Datalake.
- Tecnologías y servicios relacionados a la arquitectura moderna de datos.
- Arquitecturas de referencia.
- **Taller: Desarrollo de un diagrama de una arquitectura moderna de datos.**

#### Data Modeler Fundamentals

- Tipos de modelos de datos.
- Modelo Conceptual, físico y lógico.
- Fundamentos de normalización.
- Cardinalidad.
- OLTP vs OLAP.
- **Taller: Creación de un diagrama de Modelo lógico.**

#### Data Modeler Dimensional

- Modelos dimensionas (Star Schema y Snowflake Schema).
- Modelamiento del Data warehouse.
- Modelamiento del Data lakehouse.
- Diseño de dimensiones (SCD).
- **Taller: Diseño de un modelo OLAP a partir de uno OLTP.**

## Módulo IV: Advanced Data Warehousing (SQLSERVER, SSIS)

### Traditional Data Warehouse

- Repaso T-SQL. DML, DDL.
- Patrones de Ingesta de datos (CDC, Delta ingestion, Full ingestion).
- Optimización del rendimiento del almacenamiento (Indexing, particionado, bucketing, design by query, clustering).
- Introducción a SQL Server Integration Services (SSIS).
- Implementación de la capa Stage en SQL Server.
- Implementación de la capa Operational Data Store (ODS) en SQL Server.
- Implementación de la capa Business Data Store (BDS) en SQL Server.
- Implementación de scripts T-SQL para poblar el Data Warehouse.
- Automatización de paquetes mediante Jobs.

### Cloud BI

- Fundamentos de Cloud Computing. Modelos principales de servicios (SaaS, PaaS, IaaS, Serverless). Modelos de costos.
  - Arquitecturas de referencia de DWH en Cloud.
  - Plataformas de almacenamiento y procesamiento multicloud. (Snowflake, Dbt).
  - **Taller: Implementación de una arquitectura DWH en Azure.**
  - **Taller: Implementación de una arquitectura DWH en GCP.**
  - **Taller: Implementación de una arquitectura DWH en AWS.**
- 

## Módulo V: Data Governance & Quality Fundamentals

### Data Governance Fundamentals

- Gobernanza de datos. Qué, por qué en la organización.
- Descripción general del modelo operativo de gobernanza de datos: Centralizado, Federado e Híbrido.
- Conceptos generales del Comité de la Dirección General. Roles y funciones: Responsable de datos, Administrador de datos, propietario de datos, custodio de datos.
- Data Lineage. Definición, técnicas y tipos.
- Data Catalog & Business Glossary. Definición, importancia, componentes, desarrollo y mantenimiento.
- Data Dictionary. Definición, importancia.
- Data Quality. Definición y dimensiones (métricas).
- **Taller: Estimación de métricas de calidad de datos con Excel para diversos Datasets.**

### Evaluación de Madurez de Gobierno de datos

- **Taller: Elección del modelo operativo para un caso uso.**
  - **Taller: Revisión de áreas de conocimiento y niveles de madurez según DMBOK (Capítulo 15, sección 1.3).**
  - **Taller: Plantilla 1. Diagnóstico de brechas (actual vs. Esperado). GAP y técnicas de priorización.**
  - **Taller: Plantilla 2. Progreso de actividades por área de conocimiento.**
  - **Taller: Plantilla 3. Identificación y análisis de Stakeholders.**
  - **Taller: Plantilla 4. Planificación de proyectos por áreas de conocimiento priorizadas.**
- 

## Módulo VI: Data Security Fundamentals

### Marco Normativo

- Overview marcos normativos GDPR, CCPA, PCI DSS.
- Overview ISO 8000 Data Quality, ISO 27001 Information Security Management.
- Gestión de riesgos.
- **Taller: Diseño de un plan de seguridad de datos básico: Identificación de vulnerabilidades, análisis de riesgos y plan de contingencia.**

### Protección de datos en proyectos de datos

- Encriptación simétrica, encriptación asimétrica, enmascaramiento de datos.
  - Algoritmos comunes de encriptación.
  - Encriptación vs Tokenización
  - Servicios cloud de almacenamiento de secretos y accesos de forma segura.
  - **Taller: Diseño de políticas de seguridad de datos.**
  - **Taller: Uso de servicios cloud para almacenamiento de las credenciales.**
- 

## Módulo VII: Proyecto de Ingeniería de Datos I

### Sesión 1. Lineamientos generales

- Presentación de formatos.
- Lineamientos, sugerencias, buenas prácticas.

### Sesión 2. Presentación de primer avance. Revisión y feedback

- (I) Diseño de Data Roadmap Corporativo.
- (II) Diseño del Data Warehouse.

### Sesión 3. Presentación de segundo avance. Revisión y feedback

- (III) Data Warehouse implementado.

### Sesión 4. Presentación final y sustentación

- I + II + III
- (IV) Plan de gobierno de datos (iniciativas).
- (V) Plan de seguridad de datos (básico).

# Track 02: Corporative Data Lakehouding

## Módulo I: Storytelling & Communication Skills

### Storytelling

- Storytelling. Definición, importancia en los negocios y en el trabajo con proyectos.
- Principio, desarrollo y final de impacto.
- Implicación de las emociones.

### El arte de la persuasión

- Comprensión de la audiencia.
  - Modelos de procesamiento de información
  - Los llamamientos motivacionales, reglas de interacciones humanas.
  - Comunicación auténtica. Claves verbales y no verbales.
  - Manejo de objeciones y defensa de propuestas.
  - **Taller: Planteamiento y defensa de un proyecto de data science empleando técnicas de Storytelling y de persuasión comunicativa.**
- 

## Módulo II: Data Monetization

### Monetización de datos

- Definición e importancia de la monetización de datos.
- Estrategias de monetización externa.
- Estrategias de monetización interna.

### Valorización de proyectos de analítica

- Enfoques para la valorización de proyectos de analítica.
  - **Taller: Valorización basada en pilotos y experimentos.**
- 

## Módulo III: Advanced topics with Python

### Advanced Python I. Database Control

- Repaso de principales librerías para procesamiento de datos.
- ORM utilizando SQLAlchemy.
- ETL con pandas y SQLAlchemy
- Control de concurrencia, multiprocesamiento (Uso de ThreadPoolExecutor).
- **Taller: Construcción de un framework de procesamiento de datos.**
- **Taller: Extraer datos de un API e insertarlos en una base de datos relacional.**

### Advanced Python II. Data Extraction with APIs

- Fundamentos Web Scraping utilizando Selenium.
- 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.**

### Advanced Python III. Proactive Data Cleaning

- Anticipar problemas de limpieza de datos al importar datos tabulares a Pandas.
  - Anticipar problemas de limpieza de datos al trabajar con HTML, JSON.
  - Identificación de valores faltantes y valores atípicos en subconjuntos de datos.
  - Uso de visualizaciones para la identificación de valores inesperados.
  - **Taller: Limpieza e identificación de valores anómalos en un dataset.**
- 

## Módulo IV: Big Data Performance & Tuning (Apache Spark)

### Apache Spark Fundamentals

- Principales conceptos de Big Data.
- Módulos de Apache Spark.
- Dataframe API.
- PySpark Functions.
- **Taller: Desarrollo de un ETL en Pyspark.**

### Apache Spark for Tuning and Performance

- Reading Spark Query Plans.
- Reading Spark DAGs.
- Memory Management (Spark UI).
- Executor Tuning - Cores & Memory.
- Shuffle Partition Optimization.
- Data Partitioning.
- Bucketing.
- Using Caching.
- Solve Data Skew (With Salting Method ,Boadcast Joins).
- Dynamic Partition Pruning.
- **Taller: Optimización de un proceso ETL con Pyspark.**

## Módulo V: Data Engineering on Azure

### Databricks Platform

- Databricks Fundamentals.
- Explorando Delta Lake en Azure Databricks.
- Procesamiento del Data Lakehouse (Bronze, Silver and Gold).
- Creación de Data Catalog (Database, Tables).
- Integraciones de DevOps e implementación de CI/CD para Azure Databricks.
- **Taller: Poblado un Data Lakehouse.**
- **Taller: Creación de la taxonomía en el Workspace.**
- **Taller: Creación de Catalog en Databricks.**
- **Taller: Crear y programar la ejecución del Pipeline en Workflow.**

### Azure Data Engineering Services

- Principales Servicios de Azure para Ingeniería de datos. Denominación y modelo de costos asociado.
- Arquitecturas de referencia para procesamiento de Batch y Real-time en Azure.
- Data Factory y Data Flow. Características generales, casos de uso.
- Integración Data Factory con Azure Databricks.
- Introducción a los servicios Azure Blob Storage y Storage Account.
- Revisión de servicios: Azure EventHubs y IoT Hub. Características generales, ejemplos de implementación y uso.
- Azure SQL Database y CosmosDB. Descripción y características generales.
- Azure Synapse Analytics. Propósito del servicio, características generales.
- Fabric. Propósito del servicio, características generales.
- Azure Devops, descripción y características generales.
- **Taller: Automatización del Pipeline del Data Lakehouse utilizando Azure Databricks y Data Factory.**
- **Taller: Creación del Data Lakehouse en Azure Datalake Gen2.**
- **Taller: Conexión de Power BI a servicios de datos de Azure.**

---

## Módulo VI: Data Engineering on AWS

### AWS Data Engineering Services

- Principales Servicios de AWS para Ingeniería de datos.
- Arquitecturas de referencia para procesamiento de Batch y Real-time en AWS.
- Introducción a los servicios Glue y Databrew. Características generales, casos de uso.
- El servicio S3 y seguridad en Buckets.
- Amazon RDS (Relational Database Service). Propósito del servicio, características generales.
- Amazon Athena. Propósito del servicio, características generales.
- Amazon Quicksight. Propósito del servicio, características generales.
- **Taller: Implementación de un ETL Básico con Glue.**
- **Taller: Implementación de un Datalake en AWS.**
- **Taller: Creación de un bucket y carga de archivos utilizando el SDK de AWS (boto3).**

---

## Módulo VII: Data Engineering on GCP

### GCP Data Engineering Services

- Principales Servicios de GCP para Ingeniería de datos
- Arquitecturas de referencia para procesamiento de Batch y Real-time en GCP.
- Introducción al servicio CloudFusion. Características generales, casos de uso.oto3).
- El servicio Google Storage y la gestión de la seguridad.
- Revisión del servicio Cloud PubSub. Caract. generales, ejemplos de implementación y uso.
- BigQuery. Propósito del servicio, características generales.
- Cloud Funtions. Propósito del servicio, características generales.
- **Taller: Creación de la estructura de un Data lake en Cloud Storage.**
- **Taller: Implementación de un ETL Básico con CloudFusion.**
- **Taller: Creación de una arquitectura Real time en GCP.**

## Módulo VIII: Data Orchestration on Cloud (Apache Airflow)

### Apache Airflow on Cloud

- DAapache Airflow. Definición.
  - DAG (Direct-Acyclic-Graph). Definición y casos de uso.
  - Uso de Scheduler.
  - Task and Operator.
  - Bash Operator, Python Operator y Apache Spark Operators.
  - **Taller: Implementado un DAG.**
  - **Taller: Procesamiento de datos con Apache Airflow en AWS.**
  - **Taller: Procesamiento de datos con Apache Airflow en GCP.**
  - **Taller: Procesamiento de datos con Apache Airflow en Azure.**
- 

## Módulo IX: DataOps

### Fundamentos de DataOps

- DataOps. Definición y características.
- Devops vs DataOps.
- Conceptos asociados: Continuos-Delivery (CD) y Continuos Integrations (CI).
- Servicios de automatización de despliegue: Jenkins, Azure Devops, Github Actions, Gitlab CI).
- Fundamentos de Infraestructura como código. Definición y características de Terraform.

### Git and GitHub Actions

- El control de versiones. Definición y características.
  - Git. Definición, principales comandos.
  - GitHub. Definición.
  - Estrategias de desarrollo colaborativo.
  - **Taller: Aplicando Gitflow a nuestro proyecto en Github y Desplegando nuestro código en un servidor cloud.**
- 

## Módulo X: Proyecto de Ingeniería de Datos II

### Sesión 1. Presentación de primer avance. Revisión y feedback

- (I) Solución Scraping.
- (II) Procesamiento Spark.
- (III.a) Solución Azure.

### Sesión 2. Presentación de segundo avance. Revisión y feedback

- (III.b) Solución AWS.
- (III.c) Solución GCP.

### Sesión 3. Presentación de tercer avance. Revisión y feedback

- (IV) Automatización de flujos (en la nube elegida para el proyecto).
- (V) Plan de costos.

### Sesión 4. Presentación final y sustentación

- (I) + (II) + (III) + (IV) + (V)
- (VI) Publicación del proyecto en GitHub y GitFlow.

# Estos son los expertos que te guiarán en este *Máster*



**Leslie Lugo** 

Analytical Lead en  
**GOOGLE**



**Isui Guzman** 

BI Reporting Leader Data & Analytics en  
**WALMART DE MÉXICO Y CENTROAMÉRICA**



**Richard Quispe** 

Sub Gerente Adjunto Planning & Analytics GDHA en  
**BCP**



**Gustavo Rangel** 

Google Cloud Platform Specialist en  
**GOOGLE**



**Daniel Ayvar** 

Gerente de Gobierno de Datos en  
**MIBANCO**



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

# Estos son los expertos que te guiarán en este *Máster*



**Shirley Villacorta** 

Cibersecurity & Privacy Manager en  
**PwC PERÚ**



**Juan Salinas** 

Lead Data Engineer en  
**ENCORA INC.**



**Daniela Gonzalez** 

Directora Académica en  
**DERRAMA MAGISTERIAL**



**Jorge Bustamante** 

Líder de Estrategia Corporativa de Data & Analytics en  
**CREDICORP**



**Dennys Rafael** 

Gerente de Valor y Portafolio Data Analytics en  
**ALICORP**



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

# Estos son los expertos que te guiarán en este *Máster*



**Guillermo Acha** 

Manager chapter leader Analytics en  
**RIMAC SEGUROS**



**Christian Sanchez** 

Senior Data Engineer en  
**PACÍFICO SEGUROS**



**Anita Quevedo** 

Cloud Partner Engineer en  
**GOOGLE**



**Miguel García** 

Especialista en TI en  
**E2E SOLUTIONS**



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

# HERRAMIENTAS DE APRENDIZAJE



## PYTHON

Es un lenguaje de alto nivel de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código.



## GOOGLE CLOUD PLATF.

Ofrece soluciones que permiten satisfacer tus necesidades de almacenamiento.



## AZURE

Es una plataforma creada para construir, probar, desplegar y administrar aplicaciones y servicios.



## SQL SERVER

Genera soluciones de integración de datos de alto rendimiento.



## AWS

Crea, implementa y administra sitios web, aplicaciones o procesos.



## SPARK

Motor unificado de analíticas para procesar datos a gran escala.

# CERTIFICACIONES

Este máster incluye los siguientes certificados:

## 1. Por aprobación del **Máster Internacional en Data Engineering on Cloud**

Un total de 284 horas académicas



## 2. Adicional 01: Por aprobación de la Especialización en Cloud Data Engineering

Un total de 80 horas académicas

## 3. Certificación microsoft azure: Microsoft Certified: Azure Data Fundamentals



- Los Certificados se emiten a nombre de DMC Perú y en formato digital.
- La emisión de todo certificado es posterior a la finalización del Máster, previa verificación del cumplimiento de los requisitos antes mencionados.

# Jornada Internacional Data & Analytics

La Jornada Internacional de Data & Analytics de DMC es un evento para estudiantes y profesionales que busca compartir conocimientos y experiencias en ciencia de datos y analítica.

En el evento podrás:

- Fortalecer tus habilidades en análisis de datos, IA, big data y machine learning.
- Interactuar con profesionales y académicos de renombre.
- Ampliar tu red de contactos y explorar nuevas oportunidades de negocio.
- Adquirir conocimientos de vanguardia y descubrir nuevas posibilidades en el mundo de la analítica.

Fecha

Al finalizar el *Máster*

Modalidad

**VIRTUAL**



# Programa de Mentoring one to one

El Programa de Mentoring del Máster en Data Engineering on Cloud te ayuda a maximizar tu potencial y acelerar tu desarrollo profesional en el campo de la ciencia de datos.

## Beneficios:

- Orientación personalizada de expertos en la industria y la analítica.
- Asesoramiento estratégico, apoyo técnico y orientación profesional.
- Ayuda para identificar y alcanzar tus objetivos, resolver desafíos específicos y navegar el camino hacia una carrera exitosa.
- Valiosa red de contactos en la industria.
- Preparación para enfrentar los retos del futuro y posicionarte como un líder en data science.

Este programa exclusivo es una [oportunidad única](#) para que los estudiantes de data science reciban el apoyo y la guía que necesitan para tener éxito en su campo.

## Inicio

A partir del 2do track

## Sesiones

01 mensual

## Modo

01 mentor por alumno





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