

> Especialización en

CLOUD DATA ENGINEERING CON GCP



Presentación

DMC Institute presenta la “Especialización en Cloud Data Engineering con GCP”, que te permitirá dominar los principales conceptos y herramientas de esta plataforma, tanto para emplearlas en el desarrollo de soluciones de ingeniería de datos escalables, así como una preparación académica para tu próxima postulación a la Professional Data Engineer Certification.



Sobre esta Especialización

17

sesiones

68

horas
académicas

68

talleres
prácticos

06

proyectos 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.**
- **Mentoría especializada** con docentes expertos.
- Acompañamiento **constante.**



¿Porqué estudiar esta especialización?

Domina la ingeniería de datos desde en enfoque de Google Cloud Plaform (GCP).



Domina los tópicos y servicios de Google Cloud Platform (GCP) en torno a su Certificación Oficial Professional Data Engineer.



Aprende de manera práctica y secuencia l a implementar una solución robusta de ingeniería de datos en GCP, considerando:

- a) Preparación de la plataforma.
- b) Ingesta de datos
- c) Transformación y procesamiento con Dataflow
- d) Operación y optimización de data pipelines
- e) Uso de Big Query para la implementación de Data Warehouse
- f) Implementación de módulos de analítica avanzada y GenAI.



Objetivos de la especialización

- Diseña e implementa una plataforma de datos moderna en Google Cloud utilizando Cloud Storage, Dataplex y servicios de gobierno para soportar casos analíticos empresariales.
- Diseña e implementar mecanismos de ingestión batch y streaming para poblar el Data Lake corporativo siguiendo patrones de arquitectura empresarial.
- Diseña, desarrolla y despliega pipelines de procesamiento de datos batch y streaming utilizando Apache Beam y Dataflow.
- Opera, monitorea y optimiza pipelines de procesamiento de datos en Google Cloud utilizando Dataflow, Cloud Monitoring y Cloud Logging.
- Diseña, construye, optimiza y gobierna un Data Warehouse empresarial sobre BigQuery.
- Implementa soluciones de analítica avanzada, Machine Learning e Inteligencia Artificial Generativa sobre Google Cloud utilizando BigQuery ML y Vertex AI.

Objetivo Final

Diseña, implementa y opera una plataforma moderna de datos en Google Cloud, que incorpore Data Lakes, pipelines batch y streaming, procesamiento con Dataflow, Data Warehousing con BigQuery y capacidades de Machine Learning e Inteligencia Artificial Generativa mediante Vertex AI, siguiendo las mejores prácticas evaluadas en la certificación Professional Data Engineer.

¿A quién está dirigido?

1. Ingenieros de datos Junior próximos a postular a la Professional Data Engineer Certification de Google.
2. Ingenieros de datos Senior que deseen actualizar su tech -skills en el manejo de la plataforma GCP.
3. Ingenieros de sistemas, desarrollo de software, informática y afines interesados en incursionar en la carrera de ingeniería de datos, de gran demanda laboral.



¿Cuáles son los requisitos?



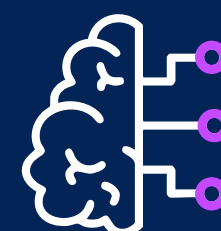
Conocimientos / Habilidades

- Conocimientos básicos de ingeniería de datos.
- Manejo básico de GCP. Recomendable contar con la certificación Google Cloud Computing Foundations.



Experiencia Laboral

- Por lo menos un año de experiencia desarrollando proyectos de ingeniería de datos en general.



Tecnológicos

- Contar con una laptop o computadora de escritorio y disponibilidad de micrófono y cámara web.
- Contar con una cuenta activa de Google Cloud Platform (GCP).

Herramientas



**Google Cloud
Platform**



Malla Curricular

I. Data Lakes & Data Platforms on GCP

1. Arquitecturas Modernas de Datos en Google Cloud

- Evolución de las arquitecturas de datos.
- Data Warehouse vs Data Lake vs Lakehouse.
- Componentes de una Modern Data Platform.
- Arquitecturas de referencia en GCP.
- Dominio "Design Data Processing Systems" del examen PDE.
- **Taller:** Creación del proyecto de laboratorio.
- **Taller:** Configuración inicial del entorno GCP.
- **Taller:** Diseño de arquitectura objetivo para el caso Retail.
- **Taller:** Identificación de fuentes, consumidores y flujos de datos.
- **Taller:** Elaboración del diagrama de arquitectura empresarial

2. Construcción del Data Lake Empresarial

- Cloud Storage como Data Lake.
- Estrategias de organización de datos.
- Capas Raw, Trusted y Curated.
- Convenciones de nombres.
- Ciclo de vida de datos.
- **Taller:** Creación de buckets corporativos.
- **Taller:** Implementación de zonas Raw, Trusted y Curated.
- **Taller:** Carga de datasets históricos.
- **Taller:** Configuración de versionado de objetos.
- **Taller:** Implementación de Lifecycle Policies.

3. Gobierno y Administración de Datos

- Principios de Data Governance.
- Metadata y Data Discovery.
- Dataplex como plataforma de gobierno.
- Data Lineage.
- Data Quality Concepts.
- **Taller:** Configuración de Dataplex.
- **Taller:** Creación de Lakes y Zones.
- **Taller:** Registro de activos de datos.
- **Taller:** Exploración de metadatos.
- **Taller:** Implementación de reglas básicas de calidad.

4. Seguridad y Administración de Accesos

- Teoría esencial.
- IAM para plataformas de datos.
- Principio de mínimo privilegio.
- Roles predefinidos y personalizados.
- Auditoría y trazabilidad.
- **Taller:** Creación de Service Accounts.
- **Taller:** Asignación de permisos por capas del Data Lake.
- **Taller:** Configuración de auditoría.
- **Taller:** Simulación de escenarios de acceso.
- **Taller:** Validación de políticas de seguridad.

5. Proyecto: Construcción de la base tecnológica de una infraestructura de ingeniería de datos

- Data Lake implementado.
- Zonas Raw, Trusted y Curated operativas.
- Gobierno inicial mediante Dataplex.
- Seguridad y control de acceso implementados.

II. Batch & Streaming Ingestion Data Pipelines

6. Fundamentos de Ingesta de Datos

- Patrones de ingestión modernos.
- ETL vs ELT.
- Batch Processing vs Stream Processing.
- Diseños desacoplados.
- Buenas prácticas de ingestión.
- **Taller:** Análisis de fuentes del caso Retail.
- **Taller:** Clasificación de cargas batch y streaming.
- **Taller:** Diseño de arquitectura de ingestión.
- **Taller:** Definición de SLA y frecuencia de carga.
- **Taller:** Diseño de estrategia incremental.

7. Ingestión Batch hacia el Data Lake

- Métodos de carga batch.
- Storage Transfer Service.
- Transferencias programadas.
- Estrategias Full Load e Incremental Load.
- **Taller:** Configuración de Storage Transfer Service.
- **Taller:** Transferencia de archivos históricos.
- **Taller:** Automatización de cargas periódicas.
- **Taller:** Validación de integridad de archivos.
- **Taller:** Implementación de particionado por fecha.

8. Ingestión Streaming con Pub/Sub

- Arquitecturas orientadas a eventos.
- Mensajería distribuida.
- Pub/Sub fundamentals.
- Entrega de mensajes y resiliencia.
- **Taller:** Creación de tópicos y suscripciones.
- **Taller:** Simulación de eventos de ventas online.
- **Taller:** Publicación de eventos en tiempo real.
- **Taller:** Consumo de mensajes.
- **Taller:** Monitoreo de throughput y latencia.

9. Calidad y Trazabilidad de Datos

- Data Quality en la ingesta.
- Validaciones estructurales y funcionales.
- Dead Letter Queues.
- Observabilidad de pipelines.
- **Taller:** Implementación de validaciones de esquema.
- **Taller:** Detección de registros inválidos.
- **Taller:** Configuración de Dead Letter Topics.
- **Taller:** Registro de métricas de calidad.
- **Taller:** Construcción de dashboard de monitoreo.

Malla Curricular

10. Proyecto: Incorporación de servicios de entrada de datos

- Cargas batch automatizadas hacia zona Raw.
- Eventos streaming ingresando al Data Lake.
- Arquitectura Pub/Sub implementada.
- Validaciones de calidad incorporadas.
- Trazabilidad y monitoreo inicial de ingestión.

III. Dataflow: Foundations & Processing Pipelines

11. Fundamentos de Apache Beam y Dataflow

- Modelo unificado de Beam.
- PCollections.
- Transformaciones.
- **Taller:** Configuración de entorno Beam.
- **Taller:** Desarrollo del primer pipeline.
- **Taller:** Ejecución local.
- **Taller:** Ejecución en Dataflow.

12. Desarrollo de Pipelines Batch

- Lectura y escritura.
- Transformaciones comunes.
- Patrones ETL.
- **Taller:** Pipeline de limpieza de ventas.
- **Taller:** Normalización de datos.
- **Taller:** Cálculo de métricas.
- **Taller:** Escritura en BigQuery.

13. Desarrollo de Pipelines Streaming

- Windowing.
- Triggers.
- Watermarks.
- **Taller:** Consumo desde Pub/Sub.
- **Taller:** Agregaciones en tiempo real.
- **Taller:** Cálculo de indicadores de negocio.
- **Taller:** Persistencia en BigQuery.

14. Buenas prácticas de desarrollo en ingeniería de datos

- Testing.
- CI/CD.
- **Taller:** Unit testing.
- **Taller:** Parameterización.
- **Taller:** Templates.
- **Taller:** Integración con repositorio Git.

15. Proyecto: Transformación y procesamiento de datos con Dataflow

- Pipelines Dataflow batch.
- Pipelines Dataflow streaming.
- Carga automática hacia BigQuery.
- Framework reutilizable para futuras integraciones.

IV. Dataflow: Operations

16. Operación de Pipelines

- Ciclo de vida de jobs.
- Escalamiento automático.
- **Taller:** Ejecución de jobs productivos.
- **Taller:** Configuración de autoscaling.
- **Taller:** Gestión de versiones.
- **Taller:** Reinicio controlado.

17. Monitoreo y Observabilidad

- Métricas operativas.
- Logging.
- Alertamiento.
- **Taller:** Cloud Monitoring.
- **Taller:** Cloud Logging.
- **Taller:** Dashboards operativos.
- **Taller:** Alertas automáticas.

18. Optimización de Costos y Rendimiento

- Optimización de recursos.
- Performance tuning.
- **Taller:** Benchmarking de pipelines.
- **Taller:** Optimización de workers.
- **Taller:** Optimización de ventanas.
- **Taller:** Evaluación de costos.

19. Troubleshooting Empresarial

- Diagnóstico de errores.
- Patrones de recuperación.
- **Taller:** Análisis de fallos reales.
- **Taller:** Recuperación de pipelines.
- **Taller:** Manejo de cuellos de botella.
- **Taller:** Simulación de incidentes.

20. Proyecto: Operación y optimización de pipelines

- Plataforma operativa monitoreada.
- Mecanismos de alerta automatizados.
- Optimización de costos.
- Operación productiva de pipelines.

V. Big Query for Data Warehouses & Knowledge Catalog

21. Diseño del Data Warehouse Empresarial

- Arquitectura analítica moderna.
- Rol de BigQuery dentro de una plataforma de datos.
- Data Marts.
- Modelado dimensional.
- Star Schema y Snowflake Schema.
- Definición de KPIs empresariales.
- **Taller:** Identificación de procesos de negocio.
- **Taller:** Diseño de modelo dimensional para Retail.
- **Taller:** Diseño de tabla de hechos de ventas.
- **Taller:** Diseño de dimensiones Cliente, Producto, Tienda y Tiempo.
- **Taller:** Elaboración del bus matrix corporativo.

Malla Curricular

22. Construcción del Data Warehouse en BigQuery

- Datasets y tablas.
- Particionamiento.
- Clustering.
- Estrategias ELT en BigQuery.
- **Taller:** Creación de datasets corporativos.
- **Taller:** Construcción de tablas de hechos.
- **Taller:** Construcción de dimensiones.
- **Taller:** Implementación de cargas ELT.
- **Taller:** Generación de capas Staging y Business.

23. Optimización y Rendimiento

- Arquitectura de ejecución de BigQuery.
- Costos y consumo de slots.
- Materialized Views.
- BI Engine.
- Buenas prácticas SQL.
- **Taller:** Optimización de consultas complejas.
- **Taller:** Implementación de particiones.
- **Taller:** Implementación de clustering.
- **Taller:** Creación de vistas materializadas.
- **Taller:** Comparación de costos antes y después de optimización.

24. Gobierno, Seguridad y Catálogo Analítico

- Gobierno de datos analíticos.
- Dataplex Catalog.
- Data Lineage.
- Row - Level Security.
- Column - Level Security.
- Data Masking.
- **Taller:** Catalogación de activos analíticos.
- **Taller:** Exploración de linaje.
- **Taller:** Implementación de Row - Level Security.
- **Taller:** Implementación de políticas de columnas sensibles.
- **Taller:** Auditoría de acceso a datasets.

25. Consumo Analítico y Preparación para IA

- Diseño de datasets para analítica avanzada.
- Feature -ready datasets.
- Preparación de datos para ML y GenAI.
- **Taller:** Construcción de tablas agregadas.
- **Taller:** Creación de datasets para predicción de ventas.
- **Taller:** Generación de métricas de comportamiento de clientes.
- **Taller:** Publicación de datasets para BigQuery ML y Vertex AI.

26. Proyecto: Transformación de datos procesados en modelos dimensionales

- Implementación del Data Warehouse empresarial.
- Implementación de Data Marts asociados.
- Habilitación de gobierno y seguridad analítica.
- Generación de algunos datasets para Machine Learning y GenAI.

VI. Big Query & GEAP (Vertex AI)

27. Introducción a IA Aplicada a Datos

- IA Generativa en Google Cloud.
- Casos de uso para Data Engineers.
- Vertex AI y ecosistema Gemini.
- **Taller:** Exploración de Vertex AI.
- **Taller:** Configuración de entornos.
- **Taller:** Primeros prompts empresariales.
- **Taller:** Integración con BigQuery.

28. BigQuery ML y Analítica Predictiva

- Machine Learning en SQL.
- Modelos supervisados.
- **Taller:** Predicción de ventas.
- **Taller:** Segmentación de clientes.
- **Taller:** Evaluación de modelos.
- **Taller:** Publicación de resultados.

29. BigQuery + Vertex AI

- Modelos fundacionales.
- Embeddings.
- Generative AI Analytics.
- **Taller:** Generación de insights automáticos.
- **Taller:** Clasificación de comentarios.
- **Taller:** Enriquecimiento semántico.
- **Taller:** Búsqueda inteligente sobre datasets.

30. Proyecto: Generación de valor para el negocio con analítica avanzada, ML y Gen AI

VII. Simulacro de Examen para Professional Data Engineer Certification

- Simulacro de Examen Oficial.
- Presentación de respuestas correctas y buenas prácticas para rendir el Examen Oficial

METODOLOGÍA LEARNING AGILE

Prepárate para nuevos retos, aprende con nuestro método de primer nivel y lánzate a conquistar el mundo.



Certificación DMC INSTITUTE

Certificado por aprobación de la Especialización en Cloud Data Engineering con Google Cloud Platform, por un total de **68 horas académicas**.

DMC | INSTITUTE



CERTIFICADO

SE OTORGA A

NOMBRES Y APELLIDOS

Por haber aprobado la Especialización en

CLOUD DATA ENGINEERING CON GOOGLE CLOUD PLATFORM

CLASES EN VIVO

Realizado del 00 de mes al 00 de mes,
completando un total de 68 horas académicas

Mes del 2026



MBA Jonny Chambi
GERENTE GENERAL



MBA Dionicio Velásquez
DIRECTOR ACADÉMICO

¿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.



Nuestros Partners

CertiProf® | Partner

Google Partners



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:





Visita nuestra web

www.dmc.pe

Síguenos en:

