

# Ingeniería de Datos en Google Cloud

Código: GOO-021

**Propuesta de Valor:** GOOGLE

**Duración:** 32 Horas



Obtenga experiencia práctica en el diseño y la construcción de sistemas de procesamiento de datos en Google Cloud. Este curso utiliza conferencias, demostraciones y laboratorios prácticos para mostrarle cómo diseñar sistemas de procesamiento de datos, crear canalizaciones de datos de extremo a extremo, analizar datos e implementar el aprendizaje automático. Este curso cubre datos estructurados, no estructurados y de transmisión.



## AUDIENCIA

Esta clase está dirigida a desarrolladores experimentados que son responsables de administrar las transformaciones de big data, que incluyen:

- Extracción, carga, transformación, limpieza y validación de datos.
- Diseño de canalizaciones y arquitecturas para el procesamiento de datos.
- Creación y mantenimiento de modelos estadísticos y de aprendizaje automático.
- Consulta de conjuntos de datos, visualización de resultados de consultas y creación de informes.



## PRE REQUISITOS

Para aprovechar al máximo este curso, los participantes deben haber:

- Completado el curso Fundamentos de Google Cloud: Big Data y aprendizaje automático o tener una experiencia equivalente.
- Competencia básica con el lenguaje de consulta común, como SQL Experience con modelado de datos, extracción, transformación, actividades de carga.
- Desarrollo de aplicaciones utilizando un lenguaje de programación común como Python.
- Familiaridad con estadísticas básicas.



## OBJETIVOS

- Diseñar y construir sistemas de procesamiento de datos en Google Cloud Platform.
- Aproveche los datos no estructurados mediante las API de Spark y ML en Cloud Dataproc.
- Procese datos por lotes y de transmisión mediante la implementación de canalizaciones de datos de ajuste de escala automático en Cloud Dataflow.
- Obtenga información empresarial a partir de conjuntos de datos extremadamente grandes con Google BigQuery.
- Entrene, evalúe y prediga con modelos de aprendizaje automático mediante TensorFlow y Cloud ML.
- Habilite información instantánea a partir de la transmisión de datos.



## CERTIFICACIÓN DISPONIBLE

- Certificación emitida por COGNOS.



## CONTENIDO

### 1. INTRODUCCIÓN A LA INGENIERÍA DE DATOS

- 1.1. EXPLORE EL ROL DE UN INGENIERO DE DATOS
- 1.2. ANALICE LOS DESAFÍOS DE LA INGENIERÍA DE DATOS
- 1.3. INTRODUCCIÓN A BIGQUERY
- 1.4. DATA LAKES Y DATA WAREHOUSES
- 1.5. DEMOSTRACIÓN: CONSULTAS FEDERADAS CON BIGQUERY
- 1.6. BASES DE DATOS TRANSACCIONALES FRENTE A ALMACENES DE DATOS
- 1.7. DEMOSTRACIÓN DEL SITIO WEB: ENCONTRAR PII EN SU CONJUNTO DE DATOS CON LA API DE DLP
- 1.8. ASÓCIESE DE MANERA EFICAZ CON OTROS EQUIPOS DE DATOS
- 1.9. GESTIONE EL ACCESO Y LA GOBERNANZA A LOS DATOS
- 1.10. CONSTRUYA TUBERÍAS LISTAS PARA PRODUCCIÓN
- 1.11. REVISE EL CASO DE ÉXITO DE UN CLIENTE DE GCP
- 1.12. LABORATORIO: ANÁLISIS DE DATOS CON BIGQUERY

### 2. CREACIÓN DE UN LAGO DE DATOS

- 2.1. INTRODUCCIÓN A DATA LAKES
- 2.2. OPCIONES DE ALMACENAMIENTO DE DATOS Y ETL EN GCP
- 2.3. CONSTRUYENDO UN LAGO DE DATOS USANDO CLOUD STORAGE
- 2.4. DEMOSTRACIÓN OPCIONAL: OPTIMIZACIÓN DE COSTOS CON CLASES DE ALMACENAMIENTO EN LA NUBE DE GOOGLE Y FUNCIONES EN LA NUBE
- 2.5. ASEGURAR EL ALMACENAMIENTO EN LA NUBE
- 2.6. ALMACENAMIENTO DE TODO TIPO DE TIPOS DE DATOS
- 2.7. DEMOSTRACIÓN DE VIDEO: EJECUCIÓN DE CONSULTAS FEDERADAS EN ARCHIVOS PARQUET Y ORC EN BIGQUERY
- 2.8. CLOUD SQL COMO UN LAGO DE DATOS RELACIONAL
- 2.9. LABORATORIO: CARGA DE DATOS DE TAXIS EN CLOUD SQL

### 3. CREACIÓN DE UN ALMACÉN DE DATOS

- 3.1. EL ALMACÉN DE DATOS MODERNO
- 3.2. INTRODUCCIÓN A BIGQUERY
- 3.3. DEMOSTRACIÓN: CONSULTA TB + DE DATOS EN SEGUNDOS
- 3.4. EMPEZANDO
- 3.5. CARGANDO DATOS
- 3.6. DEMOSTRACIÓN DE VIDEO: CONSULTA DE CLOUD SQL DESDE BIGQUERY
- 3.7. LABORATORIO: CARGA DE DATOS EN BIGQUERY
- 3.8. EXPLORANDO ESQUEMAS
- 3.9. DEMOSTRACIÓN: EXPLORACIÓN DE CONJUNTOS DE DATOS PÚBLICOS DE BIGQUERY CON SQL MEDIANTE INFORMATION\_SCHEMA
- 3.10. DISEÑO DE ESQUEMAS
- 3.11. CAMPOS ANIDADOS Y REPETIDOS
- 3.12. DEMOSTRACIÓN: CAMPOS ANIDADOS Y REPETIDOS EN BIGQUERY
- 3.13. LABORATORIO: TRABAJO CON DATOS JSON Y ARRAY EN BIGQUERY
- 3.14. OPTIMIZACIÓN CON PARTICIONES Y AGRUPACIONES
- 3.15. DEMOSTRACIÓN: TABLAS PARTICIONADAS Y AGRUPADAS EN BIGQUERY
- 3.16. VISTA PREVIA: TRANSFORMACIÓN DE DATOS POR LOTES Y TRANSMISIÓN

## 4. INTRODUCCIÓN A LA CREACIÓN DE CANALIZACIONES DE DATOS POR LOTES

- 4.1. EL, ELT, ETL
- 4.2. CONSIDERACIONES DE CALIDAD
- 4.3. CÓMO REALIZAR OPERACIONES EN BIGQUERY
- 4.4. DEMOSTRACIÓN: ELT PARA MEJORAR LA CALIDAD DE LOS DATOS EN BIGQUERY
- 4.5. DEFICIENCIAS
- 4.6. ETL PARA RESOLVER PROBLEMAS DE CALIDAD DE DATOS

## 5. EJECUCIÓN DE SPARK EN CLOUD DATAPROC

- 5.1. EL ECOSISTEMA DE HADOOP
- 5.2. EJECUTANDO HADOOP EN CLOUD DATAPROC
- 5.3. GCS EN LUGAR DE HDFS
- 5.4. OPTIMIZACIÓN DE DATAPROC
- 5.5. LABORATORIO: EJECUCIÓN DE TRABAJOS DE APACHE SPARK EN CLOUD DATAPROC

## 6. PROCESAMIENTO DE DATOS SIN SERVIDOR CON CLOUD DATAFLOW

- 6.1. CLOUD DATAFLOW
- 6.2. POR QUÉ LOS CLIENTES VALORAN DATAFLOW
- 6.3. CANALIZACIONES DE FLUJO DE DATOS
- 6.4. LABORATORIO: UNA CANALIZACIÓN DE FLUJO DE DATOS SIMPLE (PYTHON / JAVA)
- 6.5. LABORATORIO: MAPREDUCE EN DATAFLOW (PYTHON / JAVA)
- 6.6. LABORATORIO: ENTRADAS SECUNDARIAS (PYTHON / JAVA)
- 6.7. PLANTILLAS DE FLUJO DE DATOS
- 6.8. SQL DE FLUJO DE DATOS

## 7. GESTIONAR CANALIZACIONES DE DATOS CON CLOUD DATA FUSION Y CLOUD COMPOSER

- 7.1. CREA CANALIZACIONES DE DATOS POR LOTES DE FORMA VISUAL CON CLOUD DATA FUSION
- 7.2. COMPONENTES
- 7.3. DESCRIPCIÓN GENERAL DE LA INTERFAZ DE USUARIO
- 7.4. CONSTRUYENDO UNA TUBERÍA
- 7.5. EXPLORANDO DATOS USANDO WRANGLER
- 7.6. LABORATORIO: CREACIÓN Y EJECUCIÓN DE UN GRÁFICO DE CANALIZACIÓN EN CLOUD DATA FUSIÓN
- 7.7. ORGANIZAR EL TRABAJO ENTRE LOS SERVICIOS DE GCP CON CLOUD COMPOSER
- 7.8. ENTORNO APACHE AIRFLOW
- 7.9. DAG Y OPERADORES
- 7.10. PROGRAMACIÓN DEL FLUJO DE TRABAJO
- 7.11. DEMOSTRACIÓN LARGA OPCIONAL: CARGA DE DATOS ACTIVADA POR EVENTOS CON CLOUD COMPOSER, CLOUD FUNCTIONS, CLOUD STORAGE Y BIGQUERY
- 7.12. SEGUIMIENTO Y REGISTRO
- 7.13. LABORATORIO: INTRODUCCIÓN A CLOUD COMPOSER

## 8. INTRODUCCIÓN AL PROCESAMIENTO DE DATOS DE TRANSMISIÓN

- 8.1. PROCESAMIENTO DE DATOS DE TRANSMISIÓN

## 9. MENSAJERÍA SIN SERVIDOR CON CLOUD PUB / SUB

- 9.1. CLOUD PUB / SUB
- 9.2. LABORATORIO: PUBLICAR DATOS DE TRANSMISIÓN EN PUB / SUB

## 10. FUNCIONES DE TRANSMISIÓN DE CLOUD DATAFLOW

- 10.1. FUNCIONES DE TRANSMISIÓN DE CLOUD DATAFLOW
- 10.2. LABORATORIO: FLUJOS DE TRANSMISIÓN DE DATOS

## 11. FUNCIONES DE TRANSMISIÓN DE BIGQUERY Y BIGTABLE DE ALTO RENDIMIENTO

- 11.1. FUNCIONES DE TRANSMISIÓN DE BIGQUERY
- 11.2. LABORATORIO: STREAMING ANALYTICS Y DASHBOARDS
- 11.3. CLOUD BIGTABLE
- 11.4. LABORATORIO: TRANSMISIÓN DE CANALIZACIONES DE DATOS A BIGTABLE

## 12. FUNCIONALIDAD Y RENDIMIENTO AVANZADOS DE BIGQUERY

- 12.1. FUNCIONES DE LA VENTANA ANALÍTICA
- 12.2. UTILIZANDO CON CLÁUSULAS
- 12.3. FUNCIONES GIS
- 12.4. DEMOSTRACIÓN: MAPEO DE LOS CÓDIGOS POSTALES DE MÁS RÁPIDO CRECIMIENTO CON BIGQUERY GEOVIZ

12.5. CONSIDERACIONES DE RENDIMIENTO

12.6. LABORATORIO: OPTIMIZACIÓN DE LAS CONSULTAS DE BIGQUERY PARA MEJORAR EL RENDIMIENTO

12.7. LABORATORIO OPCIONAL: CREACIÓN DE TABLAS CON PARTICIONES POR FECHA EN BIGQUERY

## 13. INTRODUCCIÓN A LA ANALÍTICA Y LA IA

13.1. ¿QUÉ ES LA IA ?

13.2. DESDE ANÁLISIS DE DATOS AD-HOC HASTA DECISIONES BASADAS EN DATOS

13.3. OPCIONES PARA MODELOS DE AA EN GCP

## 14. API DE MODELOS DE AA PREDISEÑADOS PARA DATOS NO ESTRUCTURADOS

14.1. LOS DATOS NO ESTRUCTURADOS SON DIFÍCILES

14.2. API DE AA PARA ENRIQUECER DATOS

14.3. LABORATORIO: USO DE LA API DE NATURAL LANGUAGE PARA CLASIFICAR TEXTO NO ESTRUCTURADO

## 15. BIG DATA ANALYTICS CON CLOUD AI PLATFORM NOTEBOOKS

15.1. ¿QUÉ ES UN CUADERNO?

15.2. BIGQUERY MAGIC Y LAZOS CON PANDAS

15.3. LABORATORIO: BIGQUERY EN JUPYTER LABS EN AI PLATFORM

## 16. PIPELINES DE ML DE PRODUCCIÓN CON KUBEFLOW

16.1. FORMAS DE REALIZAR AA EN GCP

16.2. KUBEFLOW

16.3. AI HUB

16.4. LABORATORIO: EJECUCIÓN DE MODELOS DE IA EN KUBEFLOW

## 17. CREACIÓN DE MODELOS PERSONALIZADOS CON SQL EN BIGQUERY ML

17.1. BIGQUERY ML PARA LA CREACIÓN RÁPIDA DE MODELOS

17.2. DEMOSTRACIÓN: ENTRENE UN MODELO CON BIGQUERY ML PARA PREDECIR LAS TARIFAS DE LOS TAXIS EN NUEVA YORK

17.3. MODELOS COMPATIBLES

17.4. OPCIÓN DE LABORATORIO 1: PREDECIR LA DURACIÓN DEL VIAJE EN BICICLETA CON UN MODELO DE REGRESIÓN EN BQML

17.5. OPCIÓN DE LABORATORIO 2: RECOMENDACIONES DE PELÍCULAS EN BIGQUERY ML

## 18. CREACIÓN DE MODELOS PERSONALIZADOS CON CLOUD AUTOML

18.1. ¿POR QUÉ AUTO ML?

18.2. AUTO ML VISION

18.3. AUTO ML NLP

## ★ BENEFICIOS

- Al finalizar el curso, usted podrá diseñar sistemas de procesamiento de datos, crear canalizaciones de datos de extremo a extremo, analizar datos e implementar el aprendizaje automático.