



Data Integration with Cloud Data Fusion

Calle de la Basílica, 19
28020 Madrid
(34) 915 53 61 62
www.cas-training.com

**WE
ARE
CAS**



CAS Training



Duración
28 horas



Modalidad
Presencial



**Learning by
doing**



**Curso
Oficial**

Acerca de:

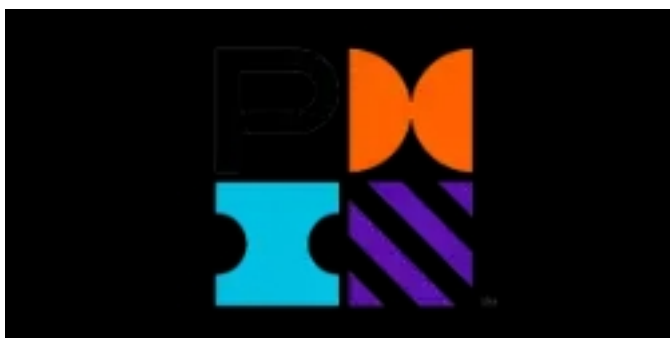
Consulta nuestro calendario Agenda tu examen de certificación

Encuentra tus cursos y certificaciones oficiales

SearchSearch



Certificación ITIL® 4 Foundation





Objetivos:

- Design and build data processing systems on Google Cloud
- Process batch and streaming data by implementing autoscaling data pipelines on Dataflow
- Derive business insights from extremely large datasets using BigQuery
- Leverage unstructured data using Spark and ML APIs on Dataproc.
- Enable instant insights from streaming data.
- Understand ML APIs and BigQuery ML, and learn to use AutoML to create powerful models without coding.

Requisitos:

Google Cloud Big Data and Machine Learning Fundamentals

Material del curso:

Documentación Oficial de Google Cloud - Data Integration with Cloud Data Fusion

Perfil del docente:

- Formador Certificado por GCP
- Más de 5 años de experiencia profesional
- Más de 4 años de experiencia docente
- Profesional activo en empresas del sector IT

Metodología:

- “Learning by doing” se centra en un contexto real y concreto, buscando un aprendizaje en equipo para la resolución de problemas en el sector empresarial.
- Aulas con grupos reducidos para que el profesional adquiera la mejor atención por parte de nuestros instructores profesionales.
- El programa de estudios como partners oficiales es confeccionado por nuestro equipo de formación y revisado por las marcas de referencia en el sector.
- La impartición de las clases podrá ser realizada tanto en modalidad Presencial como Virtual.



Contenidos:

Modulo 1: Introduction to Data Engineering

- Explore the role of a data engineer
- Analyze data engineering challenges
- Introduction to BigQuery
- Data lakes and data warehouses
- Transactional databases versus data warehouses
- Partner effectively with other data teams
- Manage data access and governance
- Manage data access and governance
- Build production-ready pipelines
- Review Google Cloud customer case study

Modulo 2: Building a Data Lake

- Introduction to data lakes
- Data storage and ETL options on Google Cloud
- Building a data lake using Cloud Storage
- Securing Cloud Storage
- Storing all sorts of data types
- Cloud SQL as a relational data lake

Modulo 3: Building a Data Warehouse

- The modern data warehouse
- Introduction to BigQuery
- Getting started with BigQuery
- Loading data
- Exploring schemas
- Schema design
- Nested and repeated fields
- Optimizing with partitioning and clustering

Modulo 4: Introduction to Building Batch Data Pipelines

- EL, ELT, ETL
- Quality considerations
- How to carry out operations in BigQuery
- Shortcomings
- ETL to solve data quality issues

Modulo 5: Executing Spark on Dataproc

- The Hadoop ecosystem



- Run Hadoop on Dataproc
- Cloud Storage instead of HDFS
- Optimize Dataproc

Modulo 6: Serverless Data Processing with Dataflow

- Introduction to Dataflow
- Why customers value Dataflow
- Dataflow pipelines
- Aggregating with GroupByKey and Combine
- Side inputs and windows
- Dataflow templates
- Dataflow SQL

Modulo 7: Manage Data Pipelines with Cloud Data Fusion and Cloud Composer

- Building batch data pipelines visually with Cloud Data Fusion
- Components
- UI overview
- Building a pipeline
- Exploring data using Wrangler
- Orchestrating work between Google Cloud services with Cloud Composer
- Apache Airflow environment
- DAGs and operators
- Workflow scheduling
- Monitoring and logging

Modulo 8: Introduction to Processing Streaming Data

- Process Streaming Data

Modulo 9: Serverless Messaging with Pub/Sub

- Introduction to Pub/Sub
- Pub/Sub push versus pull
- Publishing with Pub/Sub code

Modulo 10: Dataflow Streaming Features

- Streaming data challenges
- Dataflow windowing

Modulo 11: High-Throughput BigQuery and Bigtable Streaming Features

- Streaming into BigQuery and visualizing results
- High-throughput streaming with Cloud Bigtable
- Optimizing Cloud Bigtable performance

Modulo 12: Advanced BigQuery Functionality and Performance



- Analytic window functions
- Use With clauses
- GIS functions
- Performance considerations

Modulo 13: Introduction to Analytics and AI

- What is AI?
- From ad-hoc data analysis to data-driven decisions
- Options for ML models on Google Cloud

Modulo 14: Prebuilt ML Model APIs for Unstructured Data

- Unstructured data is hard
- ML APIs for enriching data

Modulo 15: Big Data Analytics with Notebooks

- What's a notebook?
- BigQuery magic and ties to Pandas

Modulo 16: Production ML Pipelines with Kubeflow

- Ways to do ML on Google Cloud
- ubeflow
- AI Hub

Modulo 17: Custom Model Building with SQL in BigQuery ML

- BigQuery ML for quick model building
- Supported models

Modulo 18: Custom Model Building with AutoML

- Why AutoML?
- AutoML Vision
- AutoML NLP
- AutoML tables



CAS TRAINING

UN ESPACIO PARA CRECER

cas-training.com

