

PDE: Google Professional Data Engineer

Build scalable batch and streaming data platforms on Google Cloud — and prepare for the Professional Data Engineer exam.

DURATION

4 days · 32 hours

LEVEL

Advanced

CURRICULUM

12 modules · 9 labs

DELIVERYLive online, Onsite,
Corporate

Course overview

Data engineers on Google Cloud design platforms that turn raw events and records into trusted, analysis-ready data. They choose between many services — BigQuery, Dataflow, Pub/Sub, Dataproc, Bigtable, Spanner and more — and balance latency, scale, cost and governance. This four-day course teaches those choices through hands-on pipelines and design scenarios.

Day one covers designing data processing systems: requirements, security and compliance, data migration and choosing the right storage and processing services. Day two is about ingestion and processing, with Pub/Sub, Dataflow pipelines in Apache Beam for batch and streaming, Dataproc for Spark workloads, and orchestration with Cloud Composer.

Day three focuses on storing data and preparing it for analysis: BigQuery table design, partitioning and clustering, cost controls, Bigtable and Spanner use cases, data lakes on Cloud Storage with BigLake and Dataplex, and preparing data for reporting and machine learning with BigQuery ML. Day four covers maintaining and automating workloads — monitoring, troubleshooting, scheduling, CI/CD for pipelines — and finishes with exam preparation.

Labs form one platform: events stream through Pub/Sub into Dataflow, land in BigQuery and Cloud Storage, are governed with Dataplex and analysed with SQL and BigQuery ML. This course is aligned to the Professional Data Engineer exam and prepares you for it; confirm the current exam guide on the Google Cloud certification site before booking.

Highlights

- 32 hours of live, hands-on data engineering
- Labs with BigQuery, Dataflow, Pub/Sub and Dataproc
- Batch and streaming pipelines built end to end
- Covers all Professional Data Engineer exam sections
- Design scenarios with whiteboard reviews

What's included

- End-to-end streaming and batch pipeline
- Apache Beam on Dataflow labs
- BigQuery design and cost optimisation
- Orchestration with Cloud Composer
- Governance with Dataplex and BigLake
- Scenario-based practice questions
- Course completion certificate from Ace Skills Consulting

Who this course is for

Who should attend

- Data engineers working with or moving to Google Cloud
- ETL and big data developers
- Analytics engineers and BI developers
- Data architects designing cloud data platforms

Prerequisites

- Strong SQL skills
- Working knowledge of Python or Java
- Familiarity with data warehousing and ETL concepts
- Basic Google Cloud experience recommended

Detailed curriculum — 4 days, 32 hours, 12 modules, 9 hands-on labs

Day 1 Designing Data Processing Systems

8 hours

1.1 Design foundations

- Gathering data requirements
- Choosing storage services
- Choosing processing services
- Designing for reliability and portability

1.2 Security and compliance

- IAM for data services
- Sensitive Data Protection (DLP)
- Encryption and Cloud KMS
- Data residency and sovereignty

LAB Mask sensitive fields with Sensitive Data Protection · 40 min

1.3 Data migration

- Migrating warehouses to BigQuery
- Storage Transfer Service and Transfer Appliance
- Database Migration Service
- Validating migrated data

LAB Load and validate a migrated dataset in BigQuery · 40 min

Day 2 Ingesting and Processing Data

8 hours

2.1 Messaging and ingestion

- Pub/Sub topics, subscriptions and ordering
- Batch ingestion patterns
- Datastream for change data capture

LAB Publish events to Pub/Sub and inspect subscriptions · 30 min

2.2 Dataflow and Apache Beam

- PCollections and transforms
- Templates and Flex Templates
- Windowing, triggers and late data
- Autoscaling and debugging pipelines

LAB Build a streaming Dataflow pipeline from Pub/Sub to BigQuery · 75 min

2.3 Dataproc and orchestration

- Dataproc clusters and Serverless Spark
- Workflows for lightweight orchestration
- Cloud Composer DAGs

LAB Orchestrate a batch job with Cloud Composer · 45 min

Day 3 Storing Data & Preparing It for Analysis

8 hours

3.1 BigQuery in depth

- Partitioning and clustering
- Materialised views and BI Engine
- Slots, editions and cost control
- Authorised views and row-level security

LAB Optimise a slow, expensive BigQuery workload · 50 min

3.2 Operational and specialised stores

- Bigtable schema design
- Cloud SQL and AlloyDB
- Spanner for global consistency
- Memorystore

3.3 Data lakes and analysis

- BigLake and Cloud Storage lakes
- BigQuery ML
- Dataplex governance and lineage
- Sharing data with Analytics Hub

LAB Train and evaluate a model with BigQuery ML · 40 min

Day 4 Maintaining & Automating Workloads, Exam Readiness

8 hours

4.1 Operations

- Monitoring pipelines with Cloud Monitoring
- Data quality checks
- Troubleshooting failed jobs
- Resource planning and quotas

LAB Add monitoring and alerting to the streaming pipeline · 40 min

4.2 Automation

- Scheduling queries and workflows
- Infrastructure as code for data platforms
- CI/CD for data pipelines

LAB Deploy a Dataflow template through Cloud Build · 40 min

4.3 Exam readiness

- Service-selection patterns
- Timed practice set with review
- Exam section recap

Skills and tools

Skills you will build

BigQuery

Dataflow

Apache Beam

Pub/Sub

Dataproc

Cloud Composer

Bigtable

Spanner

Dataplex

BigQuery ML

Data governance

Streaming pipelines

Tools used in class

Google Cloud console

BigQuery Studio

Apache Beam SDK

Cloud Composer (Airflow)

gcloud CLI

Cloud Build

Jupyter notebooks

Certification exam

This course is aligned to the exam below and prepares you for it. The exam itself is booked and taken with the vendor.

Exam	Google Cloud Certified – Professional Data Engineer
Exam code	PDE
Questions	50–60
Duration	120 minutes
Format	Multiple choice, Multiple response

Delivery

Online proctored, Test centre

Official page

<https://cloud.google.com/learn/certification/data-engineer>

Exam skill areas covered

Processing and storing data in systems
Managing data and data analysis loads

Exam details are pending verification. Confirm the current exam code, format and fee on the official vendor page before booking.

Frequently asked questions

Do I need to know Apache Beam?

No. Beam concepts are taught from the beginning, though Python or Java experience is needed to follow the labs.

How much machine learning is covered?

Enough for the exam: preparing data for ML and using BigQuery ML and Vertex AI at a practical level.

Is this course useful without the exam?

Yes. The pipeline you build mirrors common production designs and can serve as a reference for your own work.

How long should I prepare after the course?

Most learners revise for four to six weeks, focusing on service-selection scenarios.

Can our data team attend together?

Yes. Private batches can use sample data and pipeline patterns similar to your own.

Book a batch or talk to an advisor

Course page: aceskillsconsulting.com/course/pde-google-professional-data-engineer · Email: info@aceskillsconsulting.com

Chennai office

Guindy, Chennai, Tamil Nadu

Delhi office

B-95, Old Post Office Street, Shakarpur
New Delhi 110092
Office: 98180 12523

Google Cloud is a trademark of Google LLC. All product names, logos, brands, trademarks and registered trademarks are the property of their respective owners. Ace Skills Consulting is an independent training provider; unless explicitly stated, it is not affiliated with, endorsed by or sponsored by any vendor mentioned on this site. Exam details are provided for guidance and may change — always confirm on the vendor's official website. Curriculum generated on 23 Oct 2026; topics and order may be adjusted by the trainer to suit each batch.