

GCP Data Engineering Masterclass

BigQuery-first engineering, with real Beam pipelines on Dataflow

BigQuery Dataflow Dataproc Pub/Sub Composer

Trainer Venu Katragadda · Live online · Recordings for life

₹26,000

COURSE SYLLABUS

GCP Data Engineering Masterclass

BigQuery-first data engineering on Google Cloud, with real Beam pipelines and a Composer-orchestrated platform you build yourself.

100 hours LIVE INSTRUCTION	10 modules STRUCTURED PATH	3 projects PORTFOLIO READY	₹26,000 EMI AVAILABLE
-------------------------------	-------------------------------	-------------------------------	--------------------------

Level	Beginner to Advanced
Mode	Live, instructor-led online sessions. Every session recorded; lifetime access.
Trainer	Venu Katragadda — 14+ years in industry (L&T Infotech, Capgemini, Brillio, Manthan Systems). 1,200+ professionals trained.
Batch timings	Morning 6:00–7:30 AM IST · Morning 7:30–9:00 AM IST · Evening 7:30–9:00 PM IST · Weekend (Sat & Sun)
Certification	Google Cloud Professional Data Engineer
Enquiries	+91-8500002025 / +91-9247159150 (WhatsApp) · info@databrickstraining.com

What you will be able to do

Every outcome below is demonstrated in a lab or a graded project — not just discussed in a slide.

- Design and tune BigQuery: partitioning, clustering, slots, BI Engine, materialised views, cost controls
- Write Apache Beam pipelines and run them on Dataflow in batch and streaming
- Run Spark on Dataproc and Dataproc Serverless, including Iceberg/Delta on GCS
- Build event pipelines with Pub/Sub, Datastream CDC and Dataflow templates
- Orchestrate with Cloud Composer (Airflow) and govern with Dataplex + IAM
- Pass the Google Cloud Professional Data Engineer exam

Who this masterclass is for

- Data engineers on or moving to GCP
- Analytics engineers who live in BigQuery
- Multi-cloud engineers adding GCP
- PDE certification candidates

Prerequisites

- SQL and basic Python
- GCP free tier account (\$300 credit) — setup in session 1

Tools and technologies

BigQuery · Cloud Storage · Dataflow · Apache Beam · Dataproc · Pub/Sub · Datastream · Cloud Composer / Airflow · Dataplex · Data Catalog · Cloud Functions · Cloud Run · Looker Studio · Vertex AI · dbt · Terraform · Kafka

Curriculum — 100 hours across 10 modules

Sessions run live in small batches. Roughly 60% of each session is hands-on lab work.

MODULE 1 GCP foundations

8 hrs

- Projects, folders, organisation; IAM roles, service accounts, workload identity
- Cloud Storage: classes, lifecycle, uniform access, signed URLs
- Networking basics, VPC Service Controls, private Google access
- gcloud CLI, Terraform for GCP, billing and budget alerts
- Lab: project bootstrap with Terraform and a budget alarm

MODULE 2 BigQuery core

14 hrs

- Architecture: Dremel, Colossus, slots, shuffle; storage vs compute pricing
- Datasets, tables, views, external tables, BigLake, BigQuery Omni overview
- Partitioning (time/integer/ingestion) and clustering — with measured impact
- Loading data: batch load, Storage Write API, streaming inserts, transfers
- Standard SQL depth: window functions, arrays, structs, UNNEST, JSON, geography
- Lab: reduce a report's bytes scanned by 94%

MODULE 3 BigQuery advanced, cost and performance

10 hrs

- Query plan explanation, slot contention, reservations vs on-demand, editions
- Materialised views, BI Engine, caching, table snapshots and clones
- Scripting, stored procedures, UDFs, remote functions, BigQuery ML basics
- Time travel, fail-safe, data retention; row/column-level security and policy tags
- Cost guardrails: custom quotas, maximum bytes billed, INFORMATION_SCHEMA monitoring
- Lab: build a slot-usage and cost dashboard from INFORMATION_SCHEMA

MODULE 4 Dataflow and Apache Beam

14 hrs

- Beam model: PCollection, PTransform, ParDo, GroupByKey, Combine, side inputs
- Batch vs streaming; event time, watermarks, windows, triggers, accumulation modes
- Python and Java SDK basics; runners; Dataflow Prime and streaming engine
- Templates (classic and Flex), autoscaling, shuffle service, fusion and its pitfalls
- Error handling: dead-letter patterns, retries, update/drain of streaming jobs
- Lab: Project 2 streaming pipeline with late data and a DLQ

MODULE 5 Pub/Sub, Kafka and event ingestion

8 hrs

- Pub/Sub: topics, subscriptions (pull/push), ordering keys, exactly-once delivery, dead letters
- Pub/Sub Lite, BigQuery subscriptions, Cloud Storage subscriptions
- Kafka fundamentals: partitions, consumer groups, offsets, compaction, schema registry
- Managed Service for Apache Kafka on GCP; Confluent on GCP; when to prefer Kafka over Pub/Sub
- Datastream for CDC from MySQL/Postgres/Oracle
- Lab: CDC pipeline with schema-change handling

MODULE 6 Dataproc and Spark on GCP

10 hrs

- Dataproc clusters vs Dataproc Serverless vs Dataproc on GKE
- Sizing, preemptible/spot workers, autoscaling policies, initialization actions
- Running PySpark jobs against GCS; connector tuning; Iceberg and Delta on GCS
- Dataproc Metastore, Hive/Spark SQL, migration from on-prem Hadoop
- Lab: run a Spark job on Dataproc Serverless and compare cost vs Dataflow

MODULE 7 Orchestration with Cloud Composer / Airflow

10 hrs

- Airflow core: DAGs, TaskFlow API, operators, sensors, XCom, pools, SLAs, backfills
- Composer 2/3: environment sizing, autoscaling, plugins, private IP
- GCP operators: BigQuery, Dataflow, Dataproc, GCS, Kubernetes
- Deferrable operators, dynamic task mapping, datasets/data-aware scheduling
- CI/CD for DAGs; testing DAGs with pytest
- Lab: orchestrate Projects 1 and 3 from one Composer environment

MODULE 8 Modelling, dbt and analytics engineering

8 hrs

- Dimensional modelling in BigQuery; nested/repeated vs flat — the real trade-offs
- dbt on BigQuery: models, refs, tests, snapshots, incremental strategies, macros
- Looker Studio and Looker overview; semantic layer thinking
- Data contracts and freshness SLAs
- Lab: dbt project with tests and docs over the gold layer

MODULE 9 Governance, security and reliability

8 hrs

- Dataplex: lakes, zones, data quality tasks, automatic discovery, catalog
- IAM deep dive for data: dataset ACLs, authorised views, policy tags, DLP
- Encryption, CMEK, VPC-SC perimeters for exfiltration control
- Monitoring, logging, error reporting; SLOs for pipelines
- Lab: implement PII masking and prove it with an access test

MODULE 10 AI on GCP, capstone and PDE prep

10 hrs

- Vertex AI for data engineers: feature store, pipelines, model endpoints
- BigQuery ML and Gemini in BigQuery: SQL-native ML and AI functions
- Vector search in BigQuery for RAG use cases
- Capstone presentation and architecture review
- Professional Data Engineer exam guide, 120 practice questions, two mocks
- GCP interview questions and resume framing

Hands-on projects

You leave with work you can demo in an interview. Each project is code-reviewed and returned with written feedback.

PROJECT 1 BigQuery analytics platform

Ingest, model and serve a 2 TB dataset with partitioning, clustering, materialised views and a slot-reservation cost model — with before/after cost proof.

PROJECT 2 Streaming with Beam

Pub/Sub → Dataflow streaming pipeline with windowing, triggers, side inputs and BigQuery streaming inserts; includes a replay/backfill path.

PROJECT 3 CDC to lakehouse

Datastream from Cloud SQL → GCS → Dataproc Serverless (Iceberg) → BigQuery external tables, orchestrated by Composer.

Frequently asked questions

Dataflow or Dataproc — which does the course favour?

Neither blindly. You build the same pipeline both ways in Modules 4 and 6 and produce a cost/latency comparison, so you can defend the choice in an interview or design review.

How much will GCP cost me?

The \$300 free credit covers the entire course comfortably if you follow the teardown checklist at the end of each lab. BigQuery's 1 TB/month free query tier covers most SQL work.

Do I need Java for Beam?

No. We teach Beam in Python. Java examples are shown for reading, because some Dataflow templates and older documentation are Java-only.

Is the Professional Data Engineer exam hard?

It is case-study heavy and tests judgement more than syntax. The course includes two full timed mocks and a decision-framework session specifically for the scenario questions.

How to enrol

- Book a free demo. Sit in on a live session with the current batch — no card, no commitment. WhatsApp +91-9247159150 or call +91-8500002025.
- Pick your batch. Weekday morning, weekday evening or weekend. All times IST.
- Pay the fee (₹26,000, instalments and EMI available) and get your workspace, material and calendar invites.
- Start building. Labs from session one; recordings posted the same day.

What is included

- 100 hours of live instruction by Venu Katragadda
- Lifetime access to all session recordings
- 3 code-reviewed portfolio projects
- Preparation for: Google Cloud Professional Data Engineer
- Resume review, LinkedIn optimisation and one mock interview with written feedback
- Free re-attend of any future batch of the same masterclass
- Doubt-clearing sessions and a private student group

Refund policy in brief

When you cancel	Refund
Before the batch starts	Up to 90%
Within 5 days of start, max 2 classes attended	Up to 70%
After 7 days, or more than 5 classes attended	No refund — free transfer to a later batch

Full policy: databrickstraining.com/refund-policy/

Talk to us

Call	+91-8500002025 · +91-9247159150
WhatsApp	+91-9247159150 — usually answered within the hour
Email	info@databrickstraining.com
Website	databrickstraining.com/courses/gcp-data-engineering-masterclass/
Foundation tier	databrickstraining.in — shorter 70–80 hour courses if this one is too advanced
Institute	Sreyobhilashi IT · Udyam UDYAM-TS-02-0334643 · 112, Samshraya Homes, opp. PF Office, Balanagar, Kukatpally, Hyderabad, Telangana 500037

Sreyobhilashi IT is an independent training provider and is not affiliated with, endorsed by or sponsored by Databricks, Amazon, Microsoft, Google, Snowflake, dbt Labs or Anthropic. Placement assistance is a best-effort service, not a guarantee of employment. Cloud platforms change frequently; confirm current pricing and service behaviour with the vendor's documentation.