Google Cloud Associate Data Practitioner

Course Overview

This unified curriculum provides the technical rigor required for pipeline orchestration while ensuring coverage of analyst-specific tasks like dashboarding and cost management. This roadmap is specifically structured to cover all four domains of the Associate Data Practitioner certification exam.

Duration: 5 Days(40 hours)

Audience: Junior Data Engineers, Data Analysts, and Cloud Practitioners

Goal: Master the end-to-end data lifecycle on Google Cloud: from ingestion and storage to visualization and machine learning.

Outcome: Full readiness for the Associate Data Practitioner Exam, with practical skills in BigQuery, Dataflow, Looker, and Cloud Composer.

Table of Contents

Part 1: Data Storage & Management

1. **Storage Fundamentals:** Determining the right tool (Cloud Storage vs. BigQuery vs. Cloud SQL) based on data structure (Structured, Semi-structured, Unstructured).

2. **Lifecycle Management:** Configuring Object Lifecycle Management rules to automate deletion/archiving for cost optimization.

3. **Access Control Security:** Implementing "Least Privilege" with IAM Roles, differentiating between Standard and Uniform bucket-level access.

4. **Disaster Recovery:** Understanding High Availability (HA) strategies, replication, and choosing between Regional vs. Multi-Regional buckets.

Part 2: Data Ingestion & Preparation

1. **Ingestion Methodologies:** Differentiating between ETL, ELT, and ETLT workflows.

2. **The Toolset:** Selecting the right transfer tool (Storage Transfer Service vs. Transfer Appliance vs. BigQuery Data Transfer Service).

3. **Data Cleaning:** Using **Cloud Data Fusion** for code-free transformations and **Dataprep** for quality assessment.

4. **Format Handling:** Working with CSV, JSON, Avro, and Parquet formats; understanding schema evolution.

Part 3: Analysis & Presentation

1. **BigQuery Essentials:** Defining and executing SQL queries to extract insights; managing datasets and tables.

2. **Visual Analytics with Looker:** Creating, modifying, and sharing dashboards; understanding Looker vs. Looker Studio use cases.

3. **LookML Basics:** Manipulating simple LookML parameters to modify data models (Dimensions vs. Measures).

4. **Notebooks in the Cloud:** Using Jupyter notebooks (Vertex AI Workbench/Colab Enterprise) for exploratory data analysis (EDA).

Part 4: Machine Learning for Practitioners

1. **BigQuery ML:** Using SQL to create, train, and evaluate ML models (Linear Regression, K-Means) directly inside the data warehouse.

2. **Generative AI Integration:** Calling pretrained Google LLMs using remote connections in BigQuery.

3. **The ML Workflow:** Planning standard projects: Data Collection $\rightarrow$ Training $\rightarrow$ Evaluation $\rightarrow$ Inference.

4. **AutoML & APIs:** Identifying use cases for Vision API, Natural Language API, and Vertex AI AutoML.

Part 5: Pipeline Orchestration

1. **Orchestration Strategy:** Selecting the right tool: Cloud Composer (Airflow) vs. Workflows vs. Scheduled Queries.

2. **Batch Processing:** Building serverless pipelines with **Dataflow** (Apache Beam).

3. **Event-Driven Architectures:** Ingesting streaming data from **Pub/Sub** into BigQuery; configuring Eventarc triggers.

4. **Observability:** Monitoring pipeline health using Cloud Logging and the Dataflow Job UI.