Study Plan: GCP-Based Financial Data Engineering Role (Full Coverage)

Overview

This study plan is designed to help you acquire all the technical and domain-specific skills necessary to succeed in a GCP-based financial data engineering role. It spans 8 weeks, with weekly goals, tools to learn, and mini-projects to apply concepts.

Week 1: GCP Fundamentals + Cloud Storage

Goals: - Understand core GCP services - Set up a project with IAM roles - Learn GCS (Google Cloud Storage)

Topics: - IAM roles & policies - GCP Console, Cloud Shell - Buckets, object lifecycle, versioning

Tools: - GCP Console, gsutil, Terraform (basic)

Mini Project: - Upload and version CSV files to GCS

Week 2: BigQuery + SQL Mastery

Goals: - Master BigQuery as a data warehouse - Learn SQL for large-scale analytics

Topics: - BigQuery datasets, tables, partitioning, clustering - CTEs, window functions, joins, UDFs

Tools: - BigQuery Console, bq CLI

Mini Project: - Analyze churn/subscription trends in sample SaaS dataset

---

Week 3: Apache Beam + Dataflow (Batch Pipelines)

Goals: - Understand Beam architecture - Build a basic batch processing pipeline

Topics: - PCollection, ParDo, GroupByKey, Combine, pipeline runners - Deploy Beam Python pipeline to Dataflow

Tools: - Apache Beam (Python), Dataflow UI, Terraform (Dataflow template)

Mini Project: - Read a GCS file, clean and aggregate data, write to BigQuery

---

Week 4: Pub/Sub + Dataflow (Streaming Pipelines)

Goals: - Implement real-time streaming pipeline - Work with windowing and triggers

Topics: - Pub/Sub topics and subscriptions - Beam windowing, watermarks, late data

Tools: - gcloud pubsub CLI, Dataflow UI

Mini Project: - Simulate ad clickstream data → Pub/Sub → Dataflow → BigQuery

---

Week 5: Financial Data Modeling & Business Concepts

Goals: - Learn key financial concepts and data schemas - Model real-world datasets

Topics: - Star/Snowflake schema - Subscription, ad revenue, royalties data modeling - LTV, CAC, MRR, ROI metrics

Tools: - BigQuery, dbdiagram.io, dbt (basic)

Mini Project: - Design schema for Premium + Ads revenue models

---

Week 6: Data Quality + Monitoring

Goals: - Implement checks and monitoring for data pipelines

Topics: - Great Expectations, unit testing Beam - Logging, error handling, alerts (Cloud Monitoring)

Tools: - Great Expectations, Stackdriver Logs

Mini Project: - Add quality checks + alerts to Week 4 pipeline

---

Week 7: Data Orchestration + Lifecycle

Goals: - Schedule and automate workflows - Manage full data product lifecycle

Topics: - Cloud Composer (Airflow), task dependencies, retries - Data product lifecycle: sourcing → exposure → documentation

Tools: - Cloud Composer, Airflow UI, Looker Studio (optional)

Mini Project: - Orchestrate end-to-end pipeline with Composer

Week 8: Final Capstone + Soft Skills

Goals: - Build full project + document thoroughly - Practice collaboration & presentation

Topics: - Clean code, GitHub repo structure, README writing - Present findings to business users

Capstone Project: - Build financial dashboard (churn, revenue, ad spend) - Full pipeline: Pub/Sub → Dataflow → BigQuery → Looker Studio - Include validation, logging, orchestration, schema

---

Ongoing Learning Resources

- GCP Skill Boost: https://www.cloudskillsboost.google/
- Beam Programming Guide: https://beam.apache.org/documentation/programming-guide/
- BigQuery SQL: https://cloud.google.com/bigquery/docs/reference/standard-sql/query-syntax
- dbt: https://docs.getdbt.com/
- Financial Modeling: https://corporatefinanceinstitute.com/

Let me know if you'd like a printable PDF or to customize this plan for part-time study.