

Syllabus :

Data Engineering With GCP

Topics Covered as Follows

Module 1: Introduction & Prerequisites

- Course overview
- Introduction to GCP
- Docker and docker-compose
- Running Postgres locally with Docker
- Setting up infrastructure on GCP with Terraform
- Preparing the environment for the course
- Homework

Module 2: Workflow Orchestration

- Data Lake
- Workflow orchestration
- Workflow orchestration with Mage
- Homework

Module 3: Data Warehouse

- Data Warehouse
- BigQuery

-
- Partitioning and clustering
 - BigQuery best practices
 - Internals of BigQuery
 - Integrating BigQuery with Airflow
 - BigQuery Machine Learning

Module 4: Analytics engineering

- Basics of analytics engineering
- dbt (data build tool)
- BigQuery and dbt
- Postgres and dbt
- dbt models
- Testing and documenting
- Deployment to the cloud and locally
- Visualizing the data with google data studio and metabase

Module 5: Batch processing

- Batch processing
- What is Spark
- Spark Dataframes
- Spark SQL
- Internals: GroupBy and joins

Module 6: Streaming

- Introduction to Kafka
- Schemas (avro)
- Kafka Streams

-
- Kafka Connect and KSQL

Module 7: Project

- Putting everything we learned to practice
- Week 7 and 8: working on your project
- Week 9: reviewing your peers