

---

# Data Engineering Bootcamp

## DETAILED SYLLABUS

---

### Overview

In our endeavour to build data culture and democratize Data Science learning, we are launching a Data Engineering Bootcamp with the help of resources contributed by DataTalks.Club. The online bootcamp will have a series of week-wise learning modules along with intuitive practice quizzes/challenges.

This is a community initiative, driven by experts and mentors, and you have the opportunity to attend it for free.

### Prerequisites

- Nil, anyone with a passion for learning can make it to the finish line :)

### Format

Tutors will provide learners with guided learning paths, resources and exercises to solve. The entire schedule, practical details, registration details will be put up very soon. A brief summary of the format can be found below:

- **Week-wise modules:** Week-wise learning modules will be released that would allow you to have a structured learning path
- **For real-time communication,** we will be using Discord. This medium will help learners to clear doubts on a real-time basis if they are stuck somewhere. In addition, this will also allow learners to interact with the mentors and fellow learners

- 
- **Live doubt clearing and mentorship sessions** will be organized every week based on the requirements of the learners

## Schedule

### Week #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

### Week #2 - Data ingestion

- Data Lake
- Workflow orchestration
- Setting up Airflow locally
- Ingesting data to GCP with Airflow
- Ingesting data to local Postgres with Airflow
- Moving data from AWS to GCP (Transfer service)
- Homework

### Week #3 - Data Warehouse

- Week 3: Data Warehouse
- Data Warehouse
- BigQuery
- Partitoning and clustering
- BigQuery best practices
- Internals of BigQuery
- Integrating BigQuery with Airflow
- BigQuery Machine Learning

---

## Week #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
- Visualising the data with google data studio and metabase

## Week #5 - Batch processing

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

## Week #6 - Streaming

- Introduction to Kafka
- Schemas (avro)
- Kafka Streams
- Kafka Connect and KSQL
- More details

## Week #7, 8 & 9: Project

- Putting everything we learned to practice
- Reviewing your peers