

walkthrough-01

January 23, 2024

1 Data Engineering Zoomcamp

1.1 Module-01 Walkthrough

1.1.1 Setup clarification, FAQs.

~ ellacharmed, 2024.01.21

1.2 Goals of Module-01

- docker_sql
 - setup and access database using containers
 - populate database by ingesting data NYC Taxi 2021 yellow
 - **query database** <— main focus

1.3 Concept: container (1)

- why?
 - local experiments, reproducibility, isolation, integration test (CI/CD), automation (pipelines)

1.4 Concept: container (2)

- how?
 - using config file(s), docker hub registry

1.5 Concept: container (3)

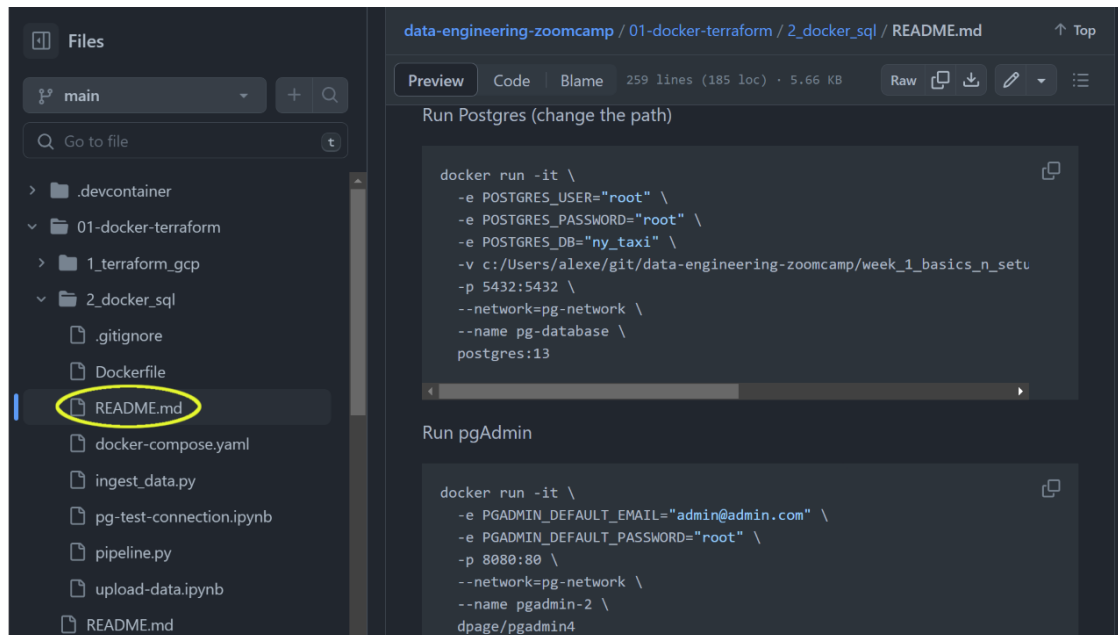
- Dockerfile vs docker-compose
 - Dockerfile : standalone, unit tests
 - docker-compose : recipe, integration

1.6 Concept: database

- server, client
- ports
- practice with postgres

1.7 Concept: db diagram

1.8 Resource: docker chapter



1.9 Goals of Module-01

- gcp_terraform
 - setup GCP account and service account
 - setup terraform IaC, with and without variables
 - learn the cycle of init-plan-apply-destroy

1.10 Concepts: google cloud platform

- Compute, Storage (bucket), BigQuery

1.11 Concepts: gcp diagram



1.12 Concepts: infrastructure

- what is IaC
- what is terraform
- the life-cycle of an infrastructure
 - init, plan, apply, destroy

1.13 Resource: terraform chapter

The screenshot displays a code editor interface. On the left, a file explorer shows a directory structure with files like `1_terraform_overview.md`, `2_gcp_overview.md`, and `README.md`. The `README.md` file is highlighted. On the right, the code preview shows the content of `README.md`, which includes sections for 'Local Setup for Terraform and GCP', 'Pre-Requisites', 'Terraform Concepts', and 'GCP setup'.

Local Setup for Terraform and GCP

Pre-Requisites

1. Terraform client installation: <https://www.terraform.io/downloads>
2. Cloud Provider account: <https://console.cloud.google.com/>

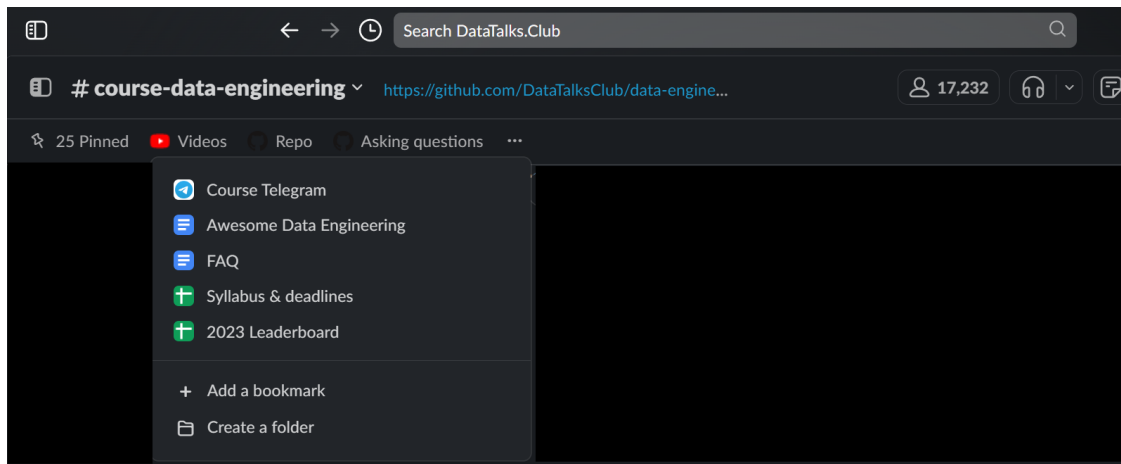
Terraform Concepts

[Terraform Overview](#)

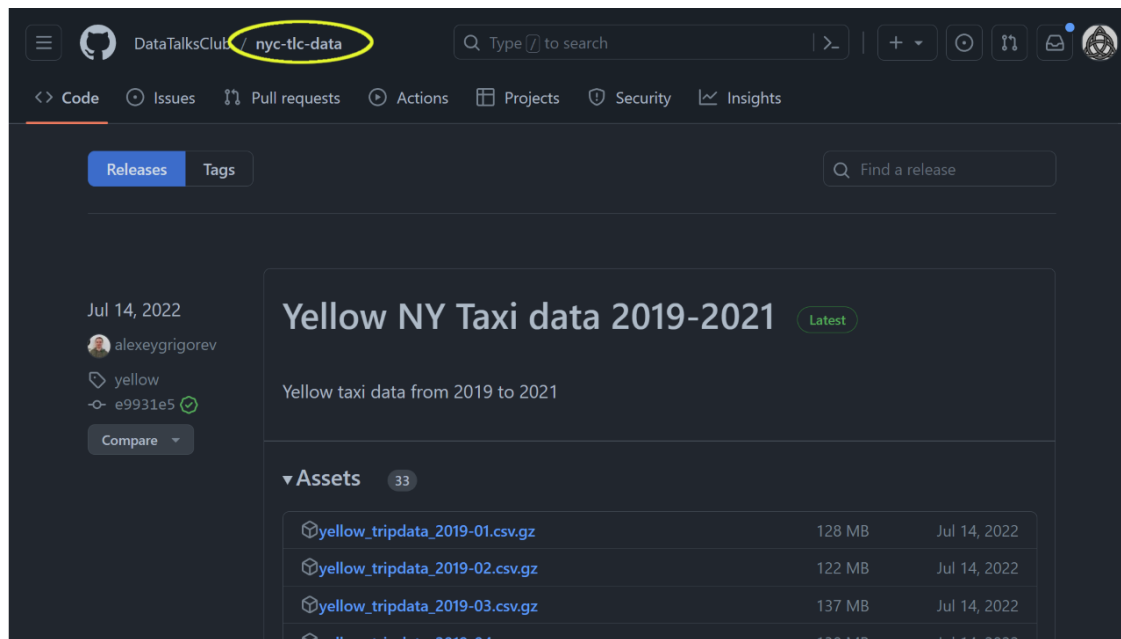
GCP setup

1. [Setup for First-time](#)
 - [Only for Windows](#) - Steps 4 & 5
2. [IAM / Access specific to this course](#)

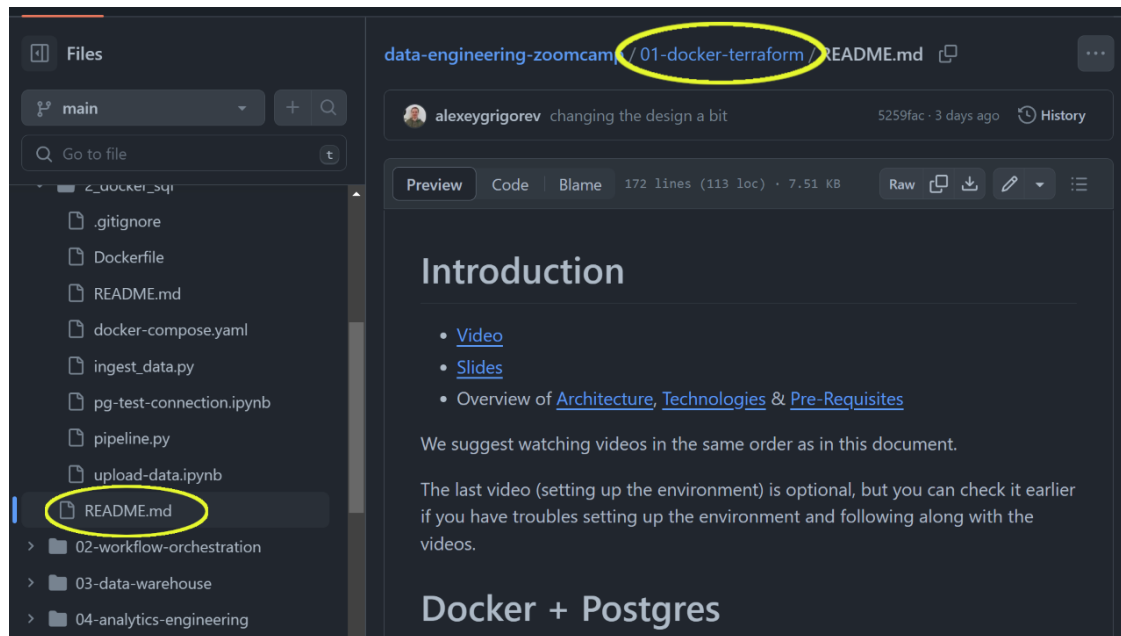
1.14 Locating resources



1.15 Resource: ny_taxi dataset



1.16 Resource: Module-01



1.17 Pre-requisite knowledge: CLI

Assumed to already know

- how to use command line (PATHs, ls, / vs , shells)
- [The Missing Semester of Your CS Education](#)
- [youtube playlist](#)
- [Scott Hanselman's My Top Tips for using Windows Terminal like a Pro](#)

1.18 Resource: missing semester of CS

The screenshot shows the YouTube playlist page for "Missing Semester IAP 2020". The left sidebar displays the playlist title, "Missing Semester", 11 videos, 312,216 views, and a last update date of Feb 2, 2020. It also shows a total duration of 10:13:46 and 11 videos counted. The main content area lists six lectures with their respective durations and view counts:

- Lecture 1: Course Overview + The Shell (2020)** - Missing Semester • 556K views • 2 years ago - 48:17
- Lecture 2: Shell Tools and Scripting (2020)** - Missing Semester • 225K views • 2 years ago - 48:56
- Lecture 3: Editors (vim) (2020)** - Missing Semester • 330K views • 2 years ago - 48:27
- Lecture 4: Data Wrangling (2020)** - Missing Semester • 130K views • 2 years ago - 50:04
- Lecture 5: Command-line Environment (2020)** - Missing Semester • 85K views • 2 years ago - 56:07
- Lecture 6: Version Control (git) (2020)** - Missing Semester • 591K views • 2 years ago

1.19 Pre-requisite knowledge: git

Assumed to already know

- how to use git and github; git clone / push / fetch / pull, remotes, branches,

1.20 Pre-requisite knowledge: venv

Assumed to already know

- how to use conda, venv, pipenv, poetry, aka environment managers etc

1.21 Pre-requisite knowledge: markdown

- how to write and read Markdown

1.22 Pre-requisite knowledge: Python

- how to write and read code in Python

1.23 Pre-requisite knowledge: networking

- concept of server and client
- concept of host and service
- watch [1.4.2 - Port Mapping and Networks in Docker \(Bonus\)](#)

1.24 Pre-requisite knowledge: docs

- how to read documentation
 - Check DATES of posts, libraries, docs last updated etc

1.25 Pre-requisite knowledge: asking questions

1.25.1 recommended skills to pick up

- GPT prompt-engineering : learn it!

1.26 Tools to install: env setup

videos: - [1.4.1 - Setting up the Environment on Google Cloud](#) - [1.4.2 - Using Github Codespaces for the Course](#) -

1.27 Tools to install: vscode

- [vs code](#)
 - useful extensions
- [docker desktop](#)

1.28 Tools to install: wsl

- [wsl2 + ubuntu](#)
 - [move_docker_distro_location](#)
 - [move_wsl_distro](#)

1.29 Tools to install: auth

- gh auth or ssh keys

1.30 Tools to install: terminal

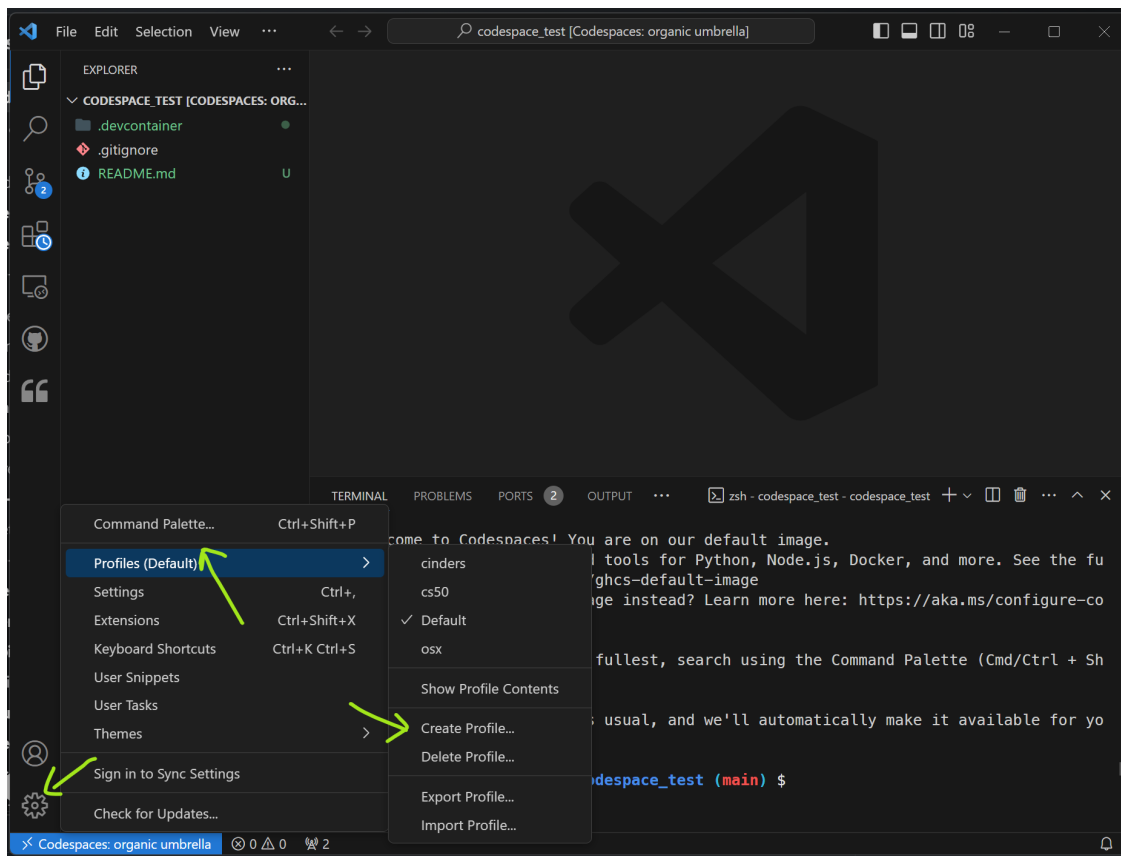
- [terminal from msft store](#)
- [hanselman's blog on console, terminal, shell](#)

1.31 Tools to install

- gcp sdk: see terraform videos
- container VMs vs local dev box
 - [micromamba env manager](#)

1.32 Tips and Tricks

Demo.



vs code extension and other tips

- outline, timeline
- create symbolic link to your documents path on windows `ln -s /mnt/c/Users/ellabelle/Github ~/projects`

1.33 FAQs

- Q: how do I start, what do I do after I sign up?
– A: See the Locate resources section

1.34 FAQs

- Q: I've watch the launch video with the Q&A, when's the next video?
– A: have you gone through absolutely all the pins and bookmarks and videos in the playlist? There's so much content

1.35 FAQs

- Q: I have an error, please help!!!!111!!
– A: Urm, the sky is blue. And...I have 2 cats. Hope that cheers you up?
– A: OK, that's bad that you have an error, is that like a tumor or the flu? Hope it's not contagious.
– A: I'm sorry that you have this error. What do you want me to do with this information?

1.36 FAQs

- Q: do I need to clone the repo?
 - A: how do you feel about typing from scratch?

1.37 FAQs

- Q: pgcli / postgres error...urm what FAQ?
 - A: has none of the FAQ suggestions help? See answer to first question

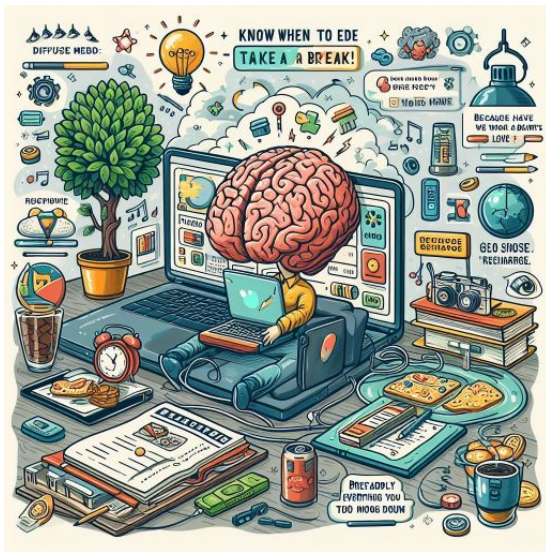
1.38 FAQs

- Q: I cannot ingest the parquet data files
 - A: See the Locate resources section

1.39 FAQs

- Q: I have error A..., then I correct that and I get error D.
 - A: OK, tell me what what you typed, what OS? What were you expecting to happen that did not happen?

1.40 Time for a break



1.41 THANK YOU!

Slides available in my repo