

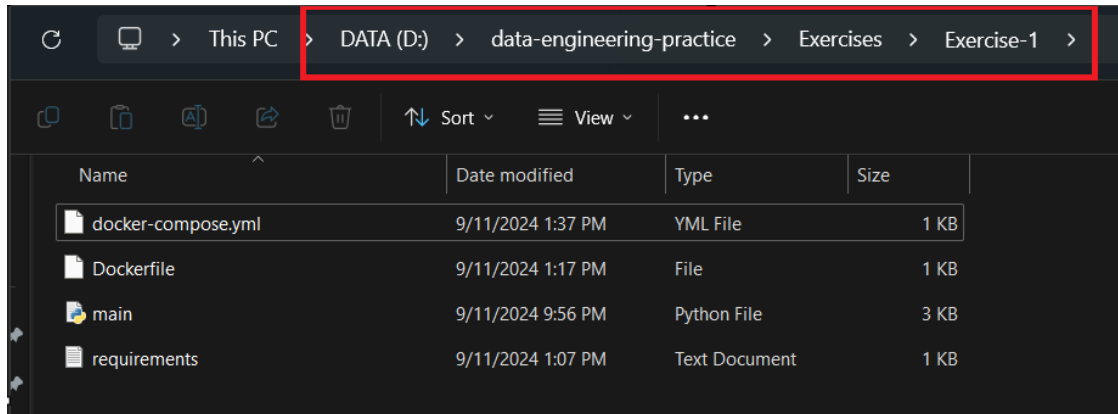
Introduction:

This guide outlines the process of downloading files using a Docker container and a Python script. The workflow utilizes Docker to manage the environment, with the files being downloaded through a multithreading approach. The entire operation, from container initialization to file extraction, is automated within the Docker ecosystem. Each step is detailed below, along with relevant figures that illustrate key stages of the process.

Step 1: Setup Docker Files and Python Script

The Dockerfile, docker-compose file, and the main Python script are located in the `Exercises/Exercises-1` folder. As shown in **Figure 1**, the `Downloads` folder has not yet been created.

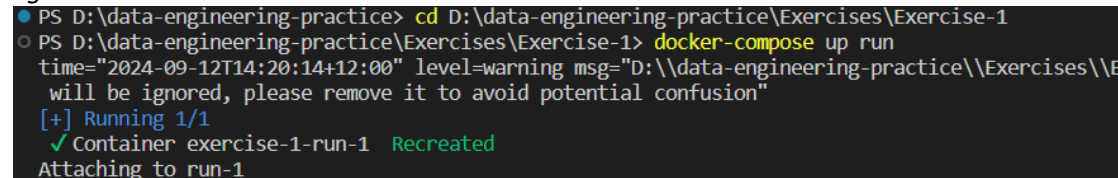
Figure 1: Initial folder structure before creating the Downloads folder.



Step 2: Starting the Docker Container

The Docker container is started using Visual Studio by running the command `docker-compose up`. This command initializes the latest image in the Docker container. **Figure 2** shows the Docker start time as `12/09/2024 - 02:20 PM`.

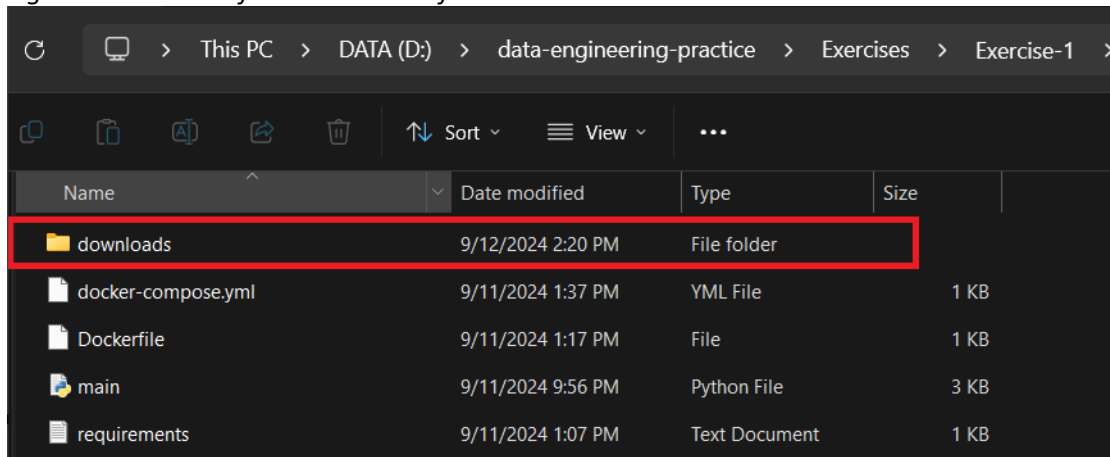
Figure 2: Docker container initialization.



Step 3: Creation of Downloads Folder

Once the Docker container is running, the Python script creates the `Downloads` folder if it does not already exist. In **Figure 3**, you can see that the folder has been created inside `Exercises/Exercises-1`.

Figure 3: Creation of the Downloads folder.



Step 4: File Download Process

The script downloads files from the provided URLs. If a file has an invalid URL, it will be skipped, while valid URLs will continue to download. **Figure 4** highlights a failed download due to an invalid URL. The download process uses a multithreading mechanism with 8 threads.

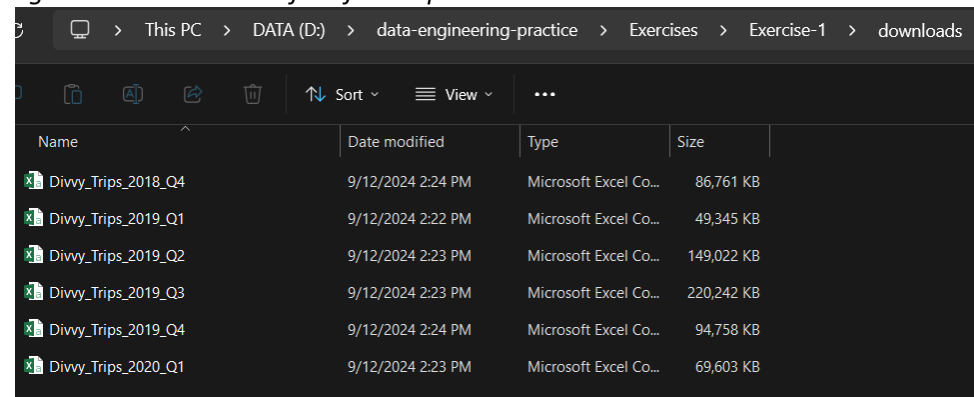
Figure 4: Handling of invalid URLs during file download.

```
Attaching to run-1
run-1 | Directory 'downloads' created.
run-1 | Failed to download: Divvy Trips 2220 Q1.zip
run-1 | Downloading 'Divvy Trips 2019 Q1.zip'...
run-1 | Downloaded 'Divvy Trips 2019 Q1.zip' to 'downloads/Divvy_Trips_2019_Q1.zip'
run-1 | Extracting Divvy_Trips_2019_Q1.csv...
run-1 | Extracted 'Divvy Trips 2019 Q1.csv' to 'downloads'
run-1 | Deleted original ZIP file 'downloads/Divvy_Trips_2019_Q1.zip'
run-1 | Downloading 'Divvy Trips 2019 Q2.zip'...
run-1 | Downloaded 'Divvy Trips 2019 Q2.zip' to 'downloads/Divvy_Trips_2019_Q2.zip'
run-1 | Extracting Divvy_Trips_2019_Q2.csv...
run-1 | Downloading 'Divvy Trips 2019 Q3.zip'...
run-1 | Downloaded 'Divvy Trips 2019 Q3.zip' to 'downloads/Divvy_Trips_2019_Q3.zip'
run-1 | Extracting Divvy_Trips_2019_Q3.csv...
run-1 | Extracted 'Divvy Trips 2019 Q2.csv' to 'downloads'
run-1 | Deleted original ZIP file 'downloads/Divvy_Trips_2019_Q2.zip'
run-1 | Extracted 'Divvy Trips 2019 Q3.csv' to 'downloads'
run-1 | Deleted original ZIP file 'downloads/Divvy_Trips_2019_Q3.zip'
run-1 | Downloading 'Divvy Trips 2020 Q1.zip'...
run-1 | Downloaded 'Divvy Trips 2020 Q1.zip' to 'downloads/Divvy_Trips_2020_Q1.zip'
run-1 | Extracting Divvy_Trips_2020_Q1.csv...
run-1 | Downloading 'Divvy Trips 2018 Q4.zip'...
run-1 | Downloaded 'Divvy Trips 2018 Q4.zip' to 'downloads/Divvy_Trips_2018_Q4.zip'
run-1 | Extracting Divvy_Trips_2018_Q4.csv...
run-1 | Extracted 'Divvy Trips 2020 Q1.csv' to 'downloads'
run-1 | Deleted original ZIP file 'downloads/Divvy_Trips_2020_Q1.zip'
run-1 | Extracted 'Divvy Trips 2018 Q4.csv' to 'downloads'
run-1 | Deleted original ZIP file 'downloads/Divvy_Trips_2018_Q4.zip'
run-1 | Downloading 'Divvy Trips 2019 Q4.zip'...
run-1 | Downloaded 'Divvy Trips 2019 Q4.zip' to 'downloads/Divvy_Trips_2019_Q4.zip'
run-1 | Extracting Divvy_Trips_2019_Q4.csv...
run-1 | Extracted 'Divvy Trips 2019 Q4.csv' to 'downloads'
run-1 | Deleted original ZIP file 'downloads/Divvy_Trips_2019_Q4.zip'
run-1 exited with code 0
```

Step 5: Extracting Files

All zip files are downloaded to the `Exercises/Exercises-1` folder. The script extracts only CSV files from the downloaded Zip files, ignoring any files in the `__MACOSX` folder. **Figure 5** shows the extracted contents.

Figure 5: Extracted CSV files from zip archives.



Name	Date modified	Type	Size
Divvy_Trips_2018_Q4	9/12/2024 2:24 PM	Microsoft Excel Co...	86,761 KB
Divvy_Trips_2019_Q1	9/12/2024 2:22 PM	Microsoft Excel Co...	49,345 KB
Divvy_Trips_2019_Q2	9/12/2024 2:23 PM	Microsoft Excel Co...	149,022 KB
Divvy_Trips_2019_Q3	9/12/2024 2:23 PM	Microsoft Excel Co...	220,242 KB
Divvy_Trips_2019_Q4	9/12/2024 2:24 PM	Microsoft Excel Co...	94,758 KB
Divvy_Trips_2020_Q1	9/12/2024 2:23 PM	Microsoft Excel Co...	69,603 KB

Step 6: Cleanup and Completion

Once all files are extracted, the script deletes the original zip files from the `Exercises/Exercises-1` folder. **Figure 6** shows the end of the Docker process, with the end time recorded as 12/09/2024 - 02:24 PM. The total process took 2 minutes.

Figure 6: Completion of the Docker process.

```
2024-09-12 14:24:40 Extracting Divvy_Trips_2019_Q4.csv...
2024-09-12 14:24:40 Extracted 'Divvy_Trips_2019_Q4.csv' to 'downloads'
2024-09-12 14:24:40 Deleted original ZIP file 'downloads/Divvy_Trips_2019_Q4.zip'
```