I think what’s missing from the module 2 videos is “How do I startup Kestra?” If you were silly like me, you found the docker-compose.yml file directly in the 02-workflow-orchestration directory and then reasoned “Ah, this looks like the multi-container application I need to start kestra and postgres in order to run flows/02\_postgres\_taxi.yml and flows/02\_postgres\_taxi\_scheduled.yml. So all I need to do is run ‘docker compose up -d’ here.” When that didn’t work at all for me , I foolishly started mucking around with changing the name of the “kestra” database to “postgres-zoomcamp” and went down a rabbithole.PRO TIP: What did work for me was to run the kestra startup command given at <https://kestra.io/docs/getting-started/quickstart> and then, changing directories to 02-workflow-orchestration, run “docker compose up -d” there. Then I can import 02-workflow/flows/02\_postgres\_taxi.yml and successfully run that flow.I admit that Will did this same “docker compose” in the video and that was the very strong hint that I should not have done the first docker-compose at all. But now, since kestra and postgres are running in separate containers, how does kestra know about postgres? There’s clearly some magic going on in that url in the pluginDefaults with host.docker.internal. This could have at least been mentioned without going into details.UPDATE: As you’ve gathered by now, I originally misunderstood what 02/docker-compose.yml is all about. It’s just for the purposes of starting Kestra. That instance of Postgres it mentions is for Kestra’s own use and has nothing to do with the postgres instance we’ll be using for New York taxis. We start that instance up in a different container by running “docker compose up -d” in 02-workflow-orchestration/postgres.To belabor the point,  I think if you want to try out executing the 02\_postgres\_taxi and 02\_postgres\_taxi\_scheduled flows, the right thing to is to run “docker compose up -d” from both directories: the 02-workflow-orchestration (start Kestra) and 02-workflow-orchestration/postgres (start postgres instance housing NY taxi data). I don’t think it matters which order you do those two runs in.I’ve struck through my original not-so-PRO “PRO TIP”, because even though it works and finding it was a sort of breakthrough for me, it’s not the best thing to do.FURTHER UPDATE: My bad — the README of 02-workflow-orchestration *does*explain the right way to start up Kestra in the “Setup Kestra” section.  These facts bespeak further strikethroughs in my original message…And the video does explain the way to start up the Postgres instance for NY taxi. Maybe there could be (redundantly, perhaps) a “Launch Postgres for NY taxi” section also in the README.  
A shout out to Will: what terrific and energetic support. (edited)

11 replies



Will RussellWill Russell  [Wednesday at 10:20 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738174853754229?thread_ts=1738174549.451149&cid=C01FABYF2RG)

It's worth noting that the command in the quickstart guide only makes a temp db in memory so it will forget all of your flows in kestra when you spin it down. There's a docker compose video in the github readme, I'd recommend checking it out: <https://youtu.be/SGL8ywf3OJQ?si=QF8yxup5G6ZCTd69>Did you check out this video too, a docker guide for this course? <https://youtu.be/73g6qJN0HcM?si=ooE06Y8JOFHZXtDX>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[Install Kestra with Docker Compose | How-to Guide](https://youtu.be/SGL8ywf3OJQ?si=QF8yxup5G6ZCTd69)

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[DE Zoomcamp FAQ - Docker Setup](https://youtu.be/73g6qJN0HcM?si=ooE06Y8JOFHZXtDX)



Will RussellWill Russell  [Wednesday at 10:28 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738175329817849?thread_ts=1738174549.451149&cid=C01FABYF2RG)

For filling the gap about host.docker.internal, here's a guide that explains it: <https://youtu.be/S1MDvdAo3oM?si=peBl8vkVW3aFugxn>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[Access Different Docker Containers on Localhost with host.docker.internal | How-to Guide](https://youtu.be/S1MDvdAo3oM?si=peBl8vkVW3aFugxn)

:+1:1



David NadlerDavid Nadler  [Wednesday at 12:13 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738181611621739?thread_ts=1738174549.451149&cid=C01FABYF2RG)

Will, Thanks for your wonderful support. I must say that I’m dazzled! I still think that it might be worth emphasizing that to follow along in the videos, one should bring up Kestra with the docker-compose,yml in 02-workflow-orchestration. The postgres database mentioned there is NOT the ny-taxi database — it’s for Kestra’s own use. Next, in an entirely  in an entirely different container, bring up an entirely different instance of postgres by running “docker compose up -d” in 02-workspace-orchestration/postgres.The best damage control I can do at this point is to edit my original message to reflect the drawbacks of using the quickstart command to start Kestra.



Will RussellWill Russell  [Wednesday at 12:20 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738182029250069?thread_ts=1738174549.451149&cid=C01FABYF2RG)

I’d recommend rewatching the start of 2.2.3 as this is all clarified there. Hope this helps: <https://youtu.be/OkfLX28Ecjg?si=EBpbgcuiNdwor9ps>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[DE Zoomcamp 2.2.3 - ETL Pipelines with Postgres in Kestra](https://youtu.be/OkfLX28Ecjg?si=EBpbgcuiNdwor9ps)



MaxkaizoMaxkaizo  [Wednesday at 5:18 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738199894955309?thread_ts=1738174549.451149&cid=C01FABYF2RG)

hey [@David Nadler](https://datatalks-club.slack.com/team/U089K4JU752) this post has been very useful to me, I was having the same problems and wasn't able to launch correctly kestra at first, so thank you.Also, now that you've commented about the readme file I checked it out and found some very useful notes of some peers that are way ahead of us in this course, I'll definitely use those in case of having problems in the future.Thanks [@Manuel Guerra](https://datatalks-club.slack.com/team/U086498R53P) and [@Horeb SEIDOU](https://datatalks-club.slack.com/team/U0846R0J3T7)

:raised_hands:2



Gregory ChoongGregory Choong  [Wednesday at 5:25 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738200338757069?thread_ts=1738174549.451149&cid=C01FABYF2RG)

Just wanted to also note that I had a difficult time as well trying to follow through.

[5:26](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738200391557579?thread_ts=1738174549.451149&cid=C01FABYF2RG)

[@Maxkaizo](https://datatalks-club.slack.com/team/U078F5ZPA5P) Thanks for reminding about the notes. Gives a much better walk through.I kept on getting stuck at running the flow in kestra since it said that it couldn't find the database

:raised_hands:2



Gregory ChoongGregory Choong  [Wednesday at 5:33 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738200826534169?thread_ts=1738174549.451149&cid=C01FABYF2RG)

reminder to modify the password though.looks like some of the notes modified it to be different than what is posted in the DE repot



Will RussellWill Russell  [Yesterday at 1:39 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738229993067489?thread_ts=1738174549.451149&cid=C01FABYF2RG)

Appreciate the feedback here!



MaxkaizoMaxkaizo  [1 hour ago](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738346262908369?thread_ts=1738174549.451149&cid=C01FABYF2RG)

I believe these videos should serve more as a walkthrough for the students of this course, without assuming any prior knowledge of the platform. For example, some of us had to discover—through trial and error—that the platform uses a PostgreSQL database.Also, there were moments where I felt overwhelmed by the number of references to other videos and materials, which made it difficult to follow along. I often had to pause, look things up, and even rewatch sections of the videos.Perhaps an introductory video outlining the overall structure of the module, with a diagram, would help provide a clearer roadmap for students.



Will RussellWill Russell  [44 minutes ago](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738346378377879?thread_ts=1738174549.451149&cid=C01FABYF2RG)

Did you check out the first video, it suggests what you should cover from our existing library first before jumping in with the exercises: <https://youtu.be/Np6QmmcgLCs>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[DE Zoomcamp 2.2.1 - Workflow Orchestration Introduction](https://youtu.be/Np6QmmcgLCs)

Hello, I have a question about module 2 of Kestra. I feel like there is no continuity with what we covered in the first week. For me it is like learning another programming language. Would it be possible to extract the *green* and *yellow* data using a Python script and then execute it from Kestra? This way, the entire logic for data extraction and loading into the database would be handled within the Python script, according to what we learn in week 1.  
I ask because, according to the videos, the explained logic seems to focus on performing everything within Kestra, leaving out the Python programming covered in week 1. Additionally, I feel that the Python workflow is more familiar to me, and I can test it as I progress.  
I appreciate your response and hope my question is clear. Best regards.

:+1:6

white_check_markeyesraised_hands

NEW

7 replies



Simone CarrieroSimone Carriero  [Wednesday at 12:28 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738139307691909?thread_ts=1738120383.864159&cid=C01FABYF2RG)

Totally agree. I had the feeling that it's focused on showing the product instead of good engineering practices. I'm watching old cohorts videos (a mix of airflow and prefect) and using dagster instead of kestra (edited)

:+1:3



Hadi MahdiHadi Mahdi  [Wednesday at 12:40 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738140010686979?thread_ts=1738120383.864159&cid=C01FABYF2RG)

Can you specify which cohort year please.



Will RussellWill Russell  [Wednesday at 12:41 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738140089115949?thread_ts=1738120383.864159&cid=C01FABYF2RG)

Running Python inside of Kestra is a common practice so if you’d prefer to go down that route, go for it. Integrating Python to Kestra is straightforward and doesn’t require many steps.<https://youtube.com/playlist?list=PLEK3H8YwZn1qMFcMmzvVws5e7UguMMND9&si=G2LcSiYxYKwFmcdQ>However it’s worth noting that the Kestra tasks are just performing SQL queries on the database, all designed to be a best practice. This allows you to focus on interacting with the DB.Also the industry use a huge variety of tools so it’s worth getting exposure to as many of them as you can! (edited)

YouTube

[Python in Kestra](https://youtube.com/playlist?list=PLEK3H8YwZn1qMFcMmzvVws5e7UguMMND9&si=G2LcSiYxYKwFmcdQ)

Share your videos with friends, family, and the world

:+1:1



Simone CarrieroSimone Carriero  [Wednesday at 1:19 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738142346191829?thread_ts=1738120383.864159&cid=C01FABYF2RG)

I see your point. What I'm trying to achieve is keeping the orchestration framework at the edge and the business logic (even just queries) decoupled from it.  
I understand that for the sake of showing how the orchestrator works it's fine also to copy/paste code into the configuration file,  
but I also think it's a good exercise to try keeping the "core logic" as decoupled as possible, maybe test it in isolation, etc

:+1:2



JENNIFER WANGJENNIFER WANG  [Wednesday at 9:38 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738172331605959?thread_ts=1738120383.864159&cid=C01FABYF2RG)

I agree as well. I'm personally using Airflow :sunglasses: .



Anna GELLERAnna GELLER  [Wednesday at 11:36 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738179381312829?thread_ts=1738120383.864159&cid=C01FABYF2RG)

A good data engineer stays open-minded and learns multiple tools. You don’t always get to pick what you use at work, so knowing both Airflow and Kestra makes you more versatile. Each has its strengths, and the best engineers understand when to use which. Expanding your toolkit only makes you better—who knows what you’ll end up using in your next role. Good luck! :rocket:

:heavy_plus_sign:2



Will RussellWill Russell  [Yesterday at 1:37 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738229837172619?thread_ts=1738120383.864159&cid=C01FABYF2RG)

To add to what Anna said, module 2 is really about understanding orchestration and seeing it in action. Understanding the concepts and where it fits into the picture is what will help you in your career!

Before starting module 2, I highly recommend watching this video from Alexey that is not on the official list but explains very well why we should use an orchestrator and why the "pipeline" of module 1 is not well designed.<https://www.youtube.com/watch?v=0yK7LXwYeD0>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [DataTalksClub ⬛](https://www.youtube.com/@DataTalksClub)

[DE Zoomcamp 2.2.1 - Introduction to Workflow Orchestration](https://www.youtube.com/watch?v=0yK7LXwYeD0)

Hey everyone. For those of you facing issues with Connection Refused inside of Kestra when connecting to your DB, we've got a new FAQ video which walks through how you can solve this with a modified docker compose file!<https://youtu.be/73g6qJN0HcM>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[DE Zoomcamp FAQ - Docker Setup](https://youtu.be/73g6qJN0HcM)

Hi everyone,I am running into an error on the dbt flow for week 2.On the earlier setups, the host.docker.internaldid not work for me, I think because I am on linux it does not work directly. I had to set up my plugin in the following way, using the name of the docker service (kestra and postgres are on the same docker-compose):

pluginDefaults:

- type: io.kestra.plugin.jdbc.postgresql

values:

url: jdbc:<postgresql://pgdatabase:5432/postgres-zoomcamp>

username: root

password: root

However, when trying to run the dbt flow from week 2, neither of these options work:

- id: dbt-build

type: io.kestra.plugin.dbt.cli.DbtCLI

env:

DBT\_DATABASE: postgres-zoomcamp

DBT\_SCHEMA: public

namespaceFiles:

enabled: true

containerImage: ghcr.io/kestra-io/dbt-postgres:latest

taskRunner:

type: io.kestra.plugin.scripts.runner.docker.Docker

commands:

- dbt deps

- "{{ inputs.dbt\_command }}"

storeManifest:

key: manifest.json

namespace: "{{ flow.namespace }}"

profiles: |

default:

outputs:

dev:

type: postgres

host: pgdatabase <-------- # Neither pgdatabase or host.docker.internal work for me!

user: root

password: root

port: 5432

dbname: postgres-zoomcamp

schema: public

threads: 8

connect\_timeout: 10

priority: interactive

target: dev

This is the error I get for both options:

[Date: 2025-01-25T09:39:09.812594Z] [Thread: MainThread] [Type: MainEncounteredError] Encountered an error:

Database Error

could not translate host name "pgdatabase" to address: Name or service not known

I've tried various things, such as creating a network on the docker compose and adding various parameters that might make host.docker.internalwork but nothing has helped.Hopefully someone can assist with this? (edited)

:+1:2:eyes:2

white_check_markeyesraised_hands

NEW

22 replies



Leonardo RodriguezLeonardo Rodriguez  [Jan 25th at 1:47 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737798470485819?thread_ts=1737798236.342679&cid=C01FABYF2RG)

[@ZoomcampQABot](https://datatalks-club.slack.com/team/U05DM3PEJA2) on week 2 host.docker.internal is not working for the dbt flow

[](https://app.slack.com/team/U05DM3PEJA2)

ZoomcampQABotZoomcampQABotAPP  [Jan 25th at 1:47 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737798478695809?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Hey, [@Leonardo Rodriguez](https://datatalks-club.slack.com/team/U031HNNSW3A)! Here you go:  
If host.docker.internal is not working for your dbt flow, consider the following potential solutions:1. \*\*Docker Network Configuration\*\*: Ensure that your Docker container is properly configured to access the host network. If you're using Docker on Windows or Mac, host…

See more

References:  
[GitHub-data-engineering-zoomcamp-03\_postgres\_dbt.yaml](https://github.com/DataTalksClub/data-engineering-zoomcamp/blob/main/02-workflow-orchestration/flows/03_postgres_dbt.yaml)  
[GitHub-data-engineering-zoomcamp-README.md](https://github.com/DataTalksClub/data-engineering-zoomcamp/blob/main/04-analytics-engineering/docker_setup/README.md)  
[FAQ-Setting up dbt locally with Docker and Postgres...](https://docs.google.com/document/d/19bnYs80DwuUimHM65UV3sylsCn2j1vziPOwzBwQrebw/edit#heading=h.ca52h8fhwrai)  
[FAQ-​​VS Code: NoPermissions (FileSystemError): Error: EACCES: permission denied (linux)...](https://docs.google.com/document/d/19bnYs80DwuUimHM65UV3sylsCn2j1vziPOwzBwQrebw/edit#heading=h.kixzcdj21bm3)

:pray: Please leave your feedback to help me improve

:thumbsup: 0

:thumbsdown: 0



Aziz KatlaneAziz Katlane  [Jan 25th at 2:58 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737802734677519?thread_ts=1737798236.342679&cid=C01FABYF2RG)

same issue here.



Leonardo RodriguezLeonardo Rodriguez  [Jan 25th at 3:39 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737805155419859?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I guess it has something to do with the tutorials being on mac or windows, while I am running on linux. There seems to be some difference.Hopefully this doesn't get buried. I've also tried this with network and other parameters, but I haven't gotten it to work:  
<https://stackoverflow.com/questions/70725881/what-is-the-equivalent-of-add-host-host-docker-internalhost-gateway-in-a-comp>

Stack OverflowStack Overflow

[What is the equivalent of --add-host=host.docker.internal:host-gateway in a Compose file](https://stackoverflow.com/questions/70725881/what-is-the-equivalent-of-add-host-host-docker-internalhost-gateway-in-a-comp)

Starting from Docker version 20.10 (<https://github.com/moby/moby/pull/40007>), there is a new special string host-gateway that one can use within the --add-host run flag to allow a direct connection...



Aziz KatlaneAziz Katlane  [Jan 25th at 3:45 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737805519888869?thread_ts=1737798236.342679&cid=C01FABYF2RG)

did you try to put the database container ip address as the host ?



Leonardo RodriguezLeonardo Rodriguez  [Jan 25th at 3:49 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737805796352409?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I haven't yet, I tried to get it working through the network and such. Did it work for you?

[3:50](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737805826742449?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I moved on at the moment, and followed with the BQ lecture. But I want to make sure I get this sorted before we get to the DBT section



Aziz KatlaneAziz Katlane  [Jan 25th at 4:30 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737808253547749?thread_ts=1737798236.342679&cid=C01FABYF2RG)

with a network already in place and when trying with the container IP , both were unsuccessful.



Leonardo RodriguezLeonardo Rodriguez  [Jan 25th at 10:49 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737830946605659?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I don't know what else to try tbh, not sure if more people are having this issue



Leonardo RodriguezLeonardo Rodriguez  [Jan 25th at 12:03 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737835419163919?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Hi [@Aziz Katlane](https://datatalks-club.slack.com/team/U06S5N3TAEQ), in case you're still looking for a solution to this. I think I've found one. It's not my preferred solution to be honest, but I couldn't solve the problem with the hostnames. That being said, at least I managed to connect dbt and postgress and was able to run the dbt flows. What I did was the following:run docker network inspect zoomcamp-network using the name you've given to your docker network and inside it you should see all containers if they are on the same network. They show up for me, which should in theory indicate that dbt can see postgres with the container name, but this still didn't work.What did work was to find at the top of the output of the inspect command for the gateway ip:

"IPAM": {

"Driver": "default",

"Options": null,

"Config": [

{

"Subnet": "172.20.0.0/16",

"Gateway": "172.20.0.1" <--- this one

}

]

I don't know if this changes from docker-compose down and up again, but I think the IPAM settings can also be set on the network parameters of the docker-compose. But I will try that tomorrow if anything. At least it's able to connect.I hope someone can shed some light on this, I am sure we are not the only ones working on linux.

:+1:1



iobrunoiobruno  [Jan 25th at 9:30 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737869425131619?thread_ts=1737798236.342679&cid=C01FABYF2RG)

[@Leonardo Rodriguez](https://datatalks-club.slack.com/team/U031HNNSW3A) Paste the docker-compose.yaml you're using here and I'll take a look later during the week



Leonardo RodriguezLeonardo Rodriguez  [Jan 26th at 1:13 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737882838833449?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Hi [@iobruno](https://datatalks-club.slack.com/team/U03SSTLJBHT), sure! thanks for taking the time to look at it. This is the docker-compose.ymlI am using. There are a few commented out lines which are other things I was trying, but that currently don't affect what made it work in the end:

networks:

network1:

name: zoomcamp-network

driver: bridge

attachable: true

# ipam:

# driver: default

# config:

# - subnet: 172.18.0.0/16

# ip\_range: 172.18.5.0/24

# gateway: 172.18.0.1

# Using named docker volumes to not have to bind specific folders and have docker manage them instead.

volumes:

postgres-data:

driver: local

kestra-data:

driver: local

zoomcamp-data:

driver: local

services:

postgres:

image: postgres

volumes:

- postgres-data:/var/lib/postgresql/data

environment:

POSTGRES\_DB: kestra

POSTGRES\_USER: kestra

POSTGRES\_PASSWORD: k3str4

healthcheck:

test: ["CMD-SHELL", "pg\_isready -d $${POSTGRES\_DB} -U $${POSTGRES\_USER}"]

interval: 30s

timeout: 10s

retries: 10

networks:

- network1

# extra\_hosts:

# - "host.docker.internal:host-gateway"

kestra:

image: kestra/kestra:latest

pull\_policy: always

# Note that this setup with a root user is intended for development purpose.

# Our base image runs without root, but the Docker Compose implementation needs root to access the Docker socket

# To run Kestra in a rootless mode in production, see: <https://kestra.io/docs/installation/podman-compose>

user: "root"

command: server standalone

volumes:

- kestra-data:/app/storage

- /var/run/docker.sock:/var/run/docker.sock

- /tmp/kestra-wd:/tmp/kestra-wd

environment:

KESTRA\_CONFIGURATION: |

datasources:

postgres:

url: jdbc:<postgresql://postgres:5432/kestra>

driverClassName: org.postgresql.Driver

username: kestra

password: k3str4

kestra:

server:

basicAuth:

enabled: false

username: "admin@kestra.io" # it must be a valid email address

password: kestra

repository:

type: postgres

storage:

type: local

local:

basePath: "/app/storage"

queue:

type: postgres

tasks:

tmpDir:

path: /tmp/kestra-wd/tmp

url: <http://localhost:8080/>

ports:

- "8080:8080"

- "8081:8081"

depends\_on:

postgres:

condition: service\_started

networks:

- network1

# extra\_hosts:

# - "host.docker.internal:host-gateway"

# Services for the local database which will have the data for the course. The other postgres is for kestra specifically

pgdatabase:

image: postgres

container\_name: pgdatabase

environment:

- POSTGRES\_USER=root

- POSTGRES\_PASSWORD=root

- POSTGRES\_DB=postgres-zoomcamp

volumes:

- zoomcamp-data:/var/lib/postgresql/data:rw # NOTE: it's not necessary to specify rw mode, but for completeness it's ok

ports:

- "5432:5432" #NOTE: Both postgres can be in the same network and port, you can use the service name to differenciate.

# extra\_hosts:

# - "host.docker.internal:host-gateway"

networks:

- network1

pgadmin:

image: dpage/pgadmin4

environment:

- PGADMIN\_DEFAULT\_EMAIL=admin@admin.com

- PGADMIN\_DEFAULT\_PASSWORD=root

ports:

- "8082:80"

networks:

- network1

# extra\_hosts:

# - "host.docker.internal:host-gateway"

My current solution was to use the gateway ip for the network where dbt asks for the postgres host. I'd like to find a more "elegant" solution, such as the host.docker.internal keyword or the name of the container as I tried to with the bridged network. Otherwise, I might set up the "IPAM" part of the network to ensure that gateway IP won't change.

:+1:2:white_check_mark:1



Jimmy HungJimmy Hung  [Jan 26th at 1:27 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737883625239069?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Hi, I had the same issue, and my solution is to find the host machine IP address and mapping it to the kestra flow host.  
On bash "ip addr show docker0", mine is 172.17.0.1, and change the dbt-build flow "host: 172.17.0.1".  
My env is WSL2 linux. (edited)

:+1:2



Leonardo RodriguezLeonardo Rodriguez  [Jan 26th at 1:50 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737885047535759?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Hi [@Jimmy Hung](https://datatalks-club.slack.com/team/U087J61J0G1), yep I've done the same basically. Though I've found the ip through inspecting the docker network. Not sure if it's exactly the same as the docker0 ip, since in theory you can set it on the docker network itself also, which should be more or less independent as you can have multiple.But the solution seems to more or less be the same, "hardcoding" the ip rather than let it figure it out like in windows or mac (edited)

:+1:1



Jimmy HungJimmy Hung  [Jan 26th at 1:57 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737885471584709?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Yep, I ran "docker run --rm alpine ping -c 4 host.docker.internal" it shows bad address, and GPT says "in some envs (non docker desktop Linux) the host.docker.internal can't map the host machine IP address, so for me I just use the IP straight :sweat_smile:



Jimmy HungJimmy Hung  [Jan 26th at 2:02 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737885779757209?thread_ts=1737798236.342679&cid=C01FABYF2RG)

<https://forums.docker.com/t/host-docker-internal-missing-in-linux-ubuntu/137535>

Docker Community ForumsDocker Community Forums

[Host.docker.internal missing in Linux (Ubuntu)](https://forums.docker.com/t/host-docker-internal-missing-in-linux-ubuntu/137535)

I’m trying to resolve an issue redirecting curl from a Docker container to the host and seeing a lot of references to host.docker.internal and host-gateway which appear to have no effect and are not in my /etc/hosts file in Ubuntu where other people indicate. Is this correct? I’m not clear whether these addresses exist for Linux as it looks like they didn’t originally but should now and if they are supported on Linux is there a way to verify they are working? OS Version Ubuntu (WSL2 on Window...

Sep 4th, 2023

:+1:1



Leonardo RodriguezLeonardo Rodriguez  [Jan 26th at 4:06 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737893193503189?thread_ts=1737798236.342679&cid=C01FABYF2RG)

yeah makes sense, though I wish it was more clear. Anyways, I'd love to just use the container name. Especially when they are all on the same "bridge" connection. But I am unsure why this is not working either.

:+1:1



CoryCory  [Jan 26th at 11:30 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737919824469419?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I also struggled on this step.  
TLDR: I got it working with the help of Leonardo and Jimmy's comments above, but I'm also looking for some help to not have to hard code redirects.  
Here is the long version of how I solved this:  
First I got the  
could not translate host name "host.docker.internal" to address: Name or service not known  error so I changed the host to "postgres" which got it to work  in flow 02 (postgres\_taxi..).  This didn't work... so I tried changing the port to 5431 like I did in my docker compose file (duh), unfortunately this didn't work either... I also tried changing host to localhost... With all of these changes I still get some version of this message (with different host / port combinations that i try):  
connection to server at "localhost" (127.0.0.1), port 5431 failed: Connection refused  
Is the server running on that host and accepting TCP/IP connections?  
Eventually i came across this thread and I added the network definition, and service components like [@Leonardo Rodriguez](https://datatalks-club.slack.com/team/U031HNNSW3A) described above. This didn't initially work on it's own it still gave me the same old are you sure the server is running message.  Reading [@Jimmy Hung](https://datatalks-club.slack.com/team/U087J61J0G1) comment above about manually redirecting I did this in my linux command line:  
ifconfig | grep docker  which gave me this output:  
docker0: flags=4099<UP,BROADCAST,MULTICAST> mtu 1500  using this I ran:  
ifconfig -a  again and scrolled down to find:  
docker0: flags=4099<UP,BROADCAST,MULTICAST> mtu 1500  
inet 172.17.0.1 netmask 255.255.0.0 broadcast 172.17.255.255  
inet6 fe80::42:51ff:fe6d:3ba2 prefixlen 64 scopeid 0x20<link>  
ether 02:42:51:6d:3b:a2 txqueuelen 0 (Ethernet)  
RX packets 2385 bytes 372058 (372.0 KB)  
RX errors 0 dropped 0 overruns 0 frame 0  
TX packets 2294 bytes 2594833 (2.5 MB)  
TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0  so I see i have the same internal redirect address that Jimmy said of 172.17.0.1.  I put this in as my host name kestra flow and executed again.. This time I got a different error:  
connection to server at "172.17.0.1", port 5431 failed: FATAL: database "postgres-zoomcamp" does not exist Now we're on to something !  I logged in to pgadmin to confirm database name which was in fact postgres. I think this came over from files used in the previous module and I didn't rename as such, but whatever this is what it is.  Now that I changed the database name as well it connects and dbt runs as expected (I think).  Here is the updated section of my kestra flow based on the above troubleshooting:  
profiles: |  
  default:  
outputs:  
dev:  
type: postgres  
#host: host.docker.internal  
host: 172.17.0.1  
#host: localhost  
user: kestra  
password: k3str4  
port: 5431  
#dbname: postgres-zoomcamp  
dbname: postgres  
schema: public  
threads: 8  
connect\_timeout: 10  
priority: interactive  
target: dev

:+1:5



Will RussellWill Russell  [Jan 27th at 2:27 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737973648526369?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I'd recommend contributing any good findings to the GitHub repo under community notes to help others discover this. I'll also add some notes to the FAQ too!



Will RussellWill Russell  [Jan 27th at 10:07 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738001263640599?thread_ts=1737798236.342679&cid=C01FABYF2RG)

Hi all, I've just spent some time adding to our FAQ page to help you with this issue: <https://github.com/DataTalksClub/data-engineering-zoomcamp/pull/628>This allows you to access containers with their names inside of Kestra

:+1:4



Albert PAlbert P  [Saturday at 7:56 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738425411124609?thread_ts=1737798236.342679&cid=C01FABYF2RG)

I used [@Will Russell](https://datatalks-club.slack.com/team/U07T39T7Y1J) docker compose file that combined 4 containers and edit the host in 03\_postgres\_dbt.yaml as below but it failed

profiles: |

default:

outputs:

dev:

type: postgres

host: postgres\_zoomcamp

Database Error  
could not translate host name "postgres\_zoomcamp" to address: Name or service not knownI placed the containers in the same network in the docker compose file as below. I change the host to 172.20.0.1 in 03\_postgres\_dbt.yaml, based on Gateway IP address from docker network inspect zoomcamp-network However it failed with a different error.

networks:

network1:

name: zoomcamp-network

driver: bridge

attachable: true

# ipam:

# driver: default

# config:

# - subnet: 172.20.0.0/16

# gateway: 172.20.0.1

volumes:

postgres-data:

driver: local

kestra-data:

driver: local

zoomcamp-data:

driver: local

services:

postgres:

image: postgres

volumes:

- postgres-data:/var/lib/postgresql/data

environment:

POSTGRES\_DB: kestra

POSTGRES\_USER: kestra

POSTGRES\_PASSWORD: k3str4

healthcheck:

test: ["CMD-SHELL", "pg\_isready -d $${POSTGRES\_DB} -U $${POSTGRES\_USER}"]

interval: 30s

timeout: 10s

retries: 10

networks:

- network1

kestra:

image: kestra/kestra:latest

pull\_policy: always

# Note that this setup with a root user is intended for development purpose.

# Our base image runs without root, but the Docker Compose implementation needs root to access the Docker socket

# To run Kestra in a rootless mode in production, see: <https://kestra.io/docs/installation/podman-compose>

user: "root"

command: server standalone

volumes:

- kestra-data:/app/storage

- /var/run/docker.sock:/var/run/docker.sock

- /tmp/kestra-wd:/tmp/kestra-wd

environment:

KESTRA\_CONFIGURATION: |

datasources:

postgres:

url: jdbc:<postgresql://postgres:5432/kestra>

driverClassName: org.postgresql.Driver

username: kestra

password: k3str4

kestra:

server:

basicAuth:

enabled: false

username: "admin@kestra.io" # it must be a valid email address

password: kestra

repository:

type: postgres

storage:

type: local

local:

basePath: "/app/storage"

queue:

type: postgres

tasks:

tmpDir:

path: /tmp/kestra-wd/tmp

url: <http://localhost:8080/>

ports:

- "8080:8080"

- "8081:8081"

depends\_on:

postgres:

condition: service\_started

networks:

- network1

postgres\_zoomcamp:

image: postgres

environment:

POSTGRES\_USER: kestra

POSTGRES\_PASSWORD: k3str4

POSTGRES\_DB: postgres-zoomcamp

ports:

- "5432:5432"

volumes:

- zoomcamp-data:/var/lib/postgresql/data

depends\_on:

kestra:

condition: service\_started

networks:

- network1

pgadmin:

image: dpage/pgadmin4

environment:

- PGADMIN\_DEFAULT\_EMAIL=admin@admin.com

- PGADMIN\_DEFAULT\_PASSWORD=root

ports:

- "8085:80"

depends\_on:

postgres\_zoomcamp:

condition: service\_started

networks:

- network1

id: 03\_postgres\_dbt

namespace: zoomcamp

inputs:

- id: dbt\_command

type: SELECT

allowCustomValue: true

defaults: dbt build

values:

- dbt build

- dbt debug # use when running the first time to validate DB connection

tasks:

- id: sync

type: io.kestra.plugin.git.SyncNamespaceFiles

url: <https://github.com/DataTalksClub/data-engineering-zoomcamp>

branch: main

namespace: "{{ flow.namespace }}"

gitDirectory: 04-analytics-engineering/taxi\_rides\_ny

dryRun: false

# disabled: true # this Git Sync is needed only when running it the first time, afterwards the task can be disabled

- id: dbt-build

type: io.kestra.plugin.dbt.cli.DbtCLI

env:

DBT\_DATABASE: postgres-zoomcamp

DBT\_SCHEMA: public

namespaceFiles:

enabled: true

containerImage: ghcr.io/kestra-io/dbt-postgres:latest

taskRunner:

type: io.kestra.plugin.scripts.runner.docker.Docker

commands:

- dbt deps

- "{{ inputs.dbt\_command }}"

storeManifest:

key: manifest.json

namespace: "{{ flow.namespace }}"

profiles: |

default:

outputs:

dev:

type: postgres

host: 172.20.0.1 #postgres\_zoomcamp dont work too

user: kestra

password: k3str4

port: 5432

dbname: postgres-zoomcamp

schema: public

threads: 8

connect\_timeout: 10

priority: interactive

target: dev

description: |

Note that you need to adjust the models/staging/schema.yml file to match your database and schema. Select and edit that Namespace File from the UI. Save and run this flow. Once <https://github.com/DataTalksClub/data-engineering-zoomcamp/pull/565/files> is merged, you can ignore this note as it will be dynamically adjusted based on env variables.

```yaml

sources:

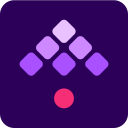
- name: staging

database: postgres-zoomcamp

schema: public

```

Database Error  
connection to server at "172.20.0.1", port 5432 failed: timeout expired (edited)

:thread:1:pushpin:1:mag:111:eyes:1



Alessandro GiacomettiAlessandro Giacometti  [Saturday at 8:47 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738428429591339?thread_ts=1737798236.342679&cid=C01FABYF2RG)

[@Albert P](https://datatalks-club.slack.com/team/U07247RRBTJ) same identical problem. Tried with network gateway IP as host in 03\_postgres\_dbt.yaml, and now the error changed to:  
"Database Error connection to server at "172.19.0.1", port 5432 failed: timeout expire"Did you find any solution? (edited)

Hey everyone. For those of you facing issues with Connection Refused inside of Kestra when connecting to your DB, we've got a new FAQ video which walks through how you can solve this with a modified docker compose file!<https://youtu.be/73g6qJN0HcM>

Hi everyone! I'm having issues trying to execute the 03\_postgres\_dbt flow. I modified its file to use host: postgres\_zoomcamp instead of host: host.docker.internal

profiles: |

default:

outputs:

dev:

type: postgres

host: postgres\_zoomcamp #instead of host.docker.internal

user: kestra

password: k3str4

port: 5432

dbname: postgres-zoomcamp

schema: public

threads: 8

connect\_timeout: 10

priority: interactive

target: dev

And I'm getting this error:  
could not translate host name "postgres\_zoomcamp" to address: Name or service not knownWhere does the dbt run? so I know how to properly connect it to the postgres db  
I'd appreciate any help, thanks! (edited)

white_check_markeyesraised_hands

NEW

3 replies



Will RussellWill Russell  [Wednesday at 10:19 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738174764096289?thread_ts=1738174054.203389&cid=C01FABYF2RG)

dbt runs inside of Kestra! :smile: Did you check out this video? <https://youtu.be/73g6qJN0HcM?si=ooE06Y8JOFHZXtDX>

[YouTube](https://www.youtube.com/)[YouTube](https://www.youtube.com/) | [Kestra](https://www.youtube.com/@kestra-io)

[DE Zoomcamp FAQ - Docker Setup](https://youtu.be/73g6qJN0HcM?si=ooE06Y8JOFHZXtDX)



Leonardo RodriguezLeonardo Rodriguez  [Wednesday at 10:30 AM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738175407573789?thread_ts=1738174054.203389&cid=C01FABYF2RG)

Hi [@Guido Carugati](https://datatalks-club.slack.com/team/U085F22MZ08), I spent a good amount of time on this issue. Hopefully this will help you. For some reason, when doing the postgres plugin on an earlier flow, the name of the docker service for postgres worked.But when trying to do it from DBT nothing worked, not even when placing them all in a bridged network and making sure they all showed up on the network.So what I did was look up the gateway IP of the docker network and I used that. Here are the details:<https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737882838833449?thread_ts=1737798236.342679&cid=C01FABYF2RG>One thing that I did to make it easier is to fill out the IPAM section for the network on the docker compose, that way the gateway ip will be what you set it to.Down below there is also a github comment by Will, though i didn't fully understand if it that docker-compose made it work on the dbt set up and why. (edited)

**Leonardo Rodriguez**

Hi [@iobruno](https://datatalks-club.slack.com/team/U03SSTLJBHT), sure! thanks for taking the time to look at it. This is the docker-compose.ymlI am using. There are a few commented out lines which are other things I was trying, but that currently don't affect what made it work in the end:

networks:

network1:

name: zoomcamp-network

Show more

[From a thread in course-data-engineering](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737882838833449?thread_ts=1737798236.342679&amp;cid=C01FABYF2RG) | [Jan 26th](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737882838833449?thread_ts=1737798236.342679&amp;cid=C01FABYF2RG) | [View reply](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1737882838833449?thread_ts=1737798236.342679&amp;cid=C01FABYF2RG)



Guido CarugatiGuido Carugati  [Wednesday at 12:55 PM](https://datatalks-club.slack.com/archives/C01FABYF2RG/p1738184112725739?thread_ts=1738174054.203389&cid=C01FABYF2RG)

Thanks, [@Leonardo Rodriguez](https://datatalks-club.slack.com/team/U031HNNSW3A)! Specifying the IP address worked well. It's probably not the best solution, but it got the job done. Thanks again!

networks: network1: name: zoomcamp-network driver: bridge attachable: true # ipam: # driver: default # config: # - subnet: 172.18.0.0/16 # ip\_range: 172.18.5.0/24 # gateway: 172.18.0.1 # Using named docker volumes to not have to bind specific folders and have docker manage them instead. volumes: postgres-data: driver: local kestra-data: driver: local zoomcamp-data: driver: local services: postgres: image: postgres volumes: - postgres-data:/var/lib/postgresql/data environment: POSTGRES\_DB: kestra POSTGRES\_USER: kestra POSTGRES\_PASSWORD: k3str4 healthcheck: test: ["CMD-SHELL", "pg\_isready -d $${POSTGRES\_DB} -U $${POSTGRES\_USER}"] interval: 30s timeout: 10s retries: 10 networks: - network1 # extra\_hosts: # - "host.docker.internal:host-gateway" kestra: image: kestra/kestra:latest pull\_policy: always # Note that this setup with a root user is intended for development purpose. # Our base image runs without root, but the Docker Compose implementation needs root to access the Docker socket # To run Kestra in a rootless mode in production, see: <https://kestra.io/docs/installation/podman-compose> user: "root" command: server standalone volumes: - kestra-data:/app/storage - /var/run/docker.sock:/var/run/docker.sock - /tmp/kestra-wd:/tmp/kestra-wd environment: KESTRA\_CONFIGURATION: | datasources: postgres: url: jdbc:<postgresql://postgres:5432/kestra> driverClassName: org.postgresql.Driver username: kestra password: k3str4 kestra: server: basicAuth: enabled: false username: "admin@kestra.io" # it must be a valid email address password: kestra repository: type: postgres storage: type: local local: basePath: "/app/storage" queue: type: postgres tasks: tmpDir: path: /tmp/kestra-wd/tmp url: <http://localhost:8080/> ports: - "8080:8080" - "8081:8081" depends\_on: postgres: condition: service\_started networks: - network1 # extra\_hosts: # - "host.docker.internal:host-gateway" # Services for the local database which will have the data for the course. The other postgres is for kestra specifically pgdatabase: image: postgres container\_name: pgdatabase environment: - POSTGRES\_USER=root - POSTGRES\_PASSWORD=root - POSTGRES\_DB=postgres-zoomcamp volumes: - zoomcamp-data:/var/lib/postgresql/data:rw # NOTE: it's not necessary to specify rw mode, but for completeness it's ok ports: - "5432:5432" #NOTE: Both postgres can be in the same network and port, you can use the service name to differenciate. # extra\_hosts: # - "host.docker.internal:host-gateway" networks: - network1 pgadmin: image: dpage/pgadmin4 environment: - PGADMIN\_DEFAULT\_EMAIL=admin@admin.com - PGADMIN\_DEFAULT\_PASSWORD=root ports: - "8082:80" networks: - network1 # extra\_hosts: # - "host.docker.internal:host-gateway"

pkill -KILL -u seed