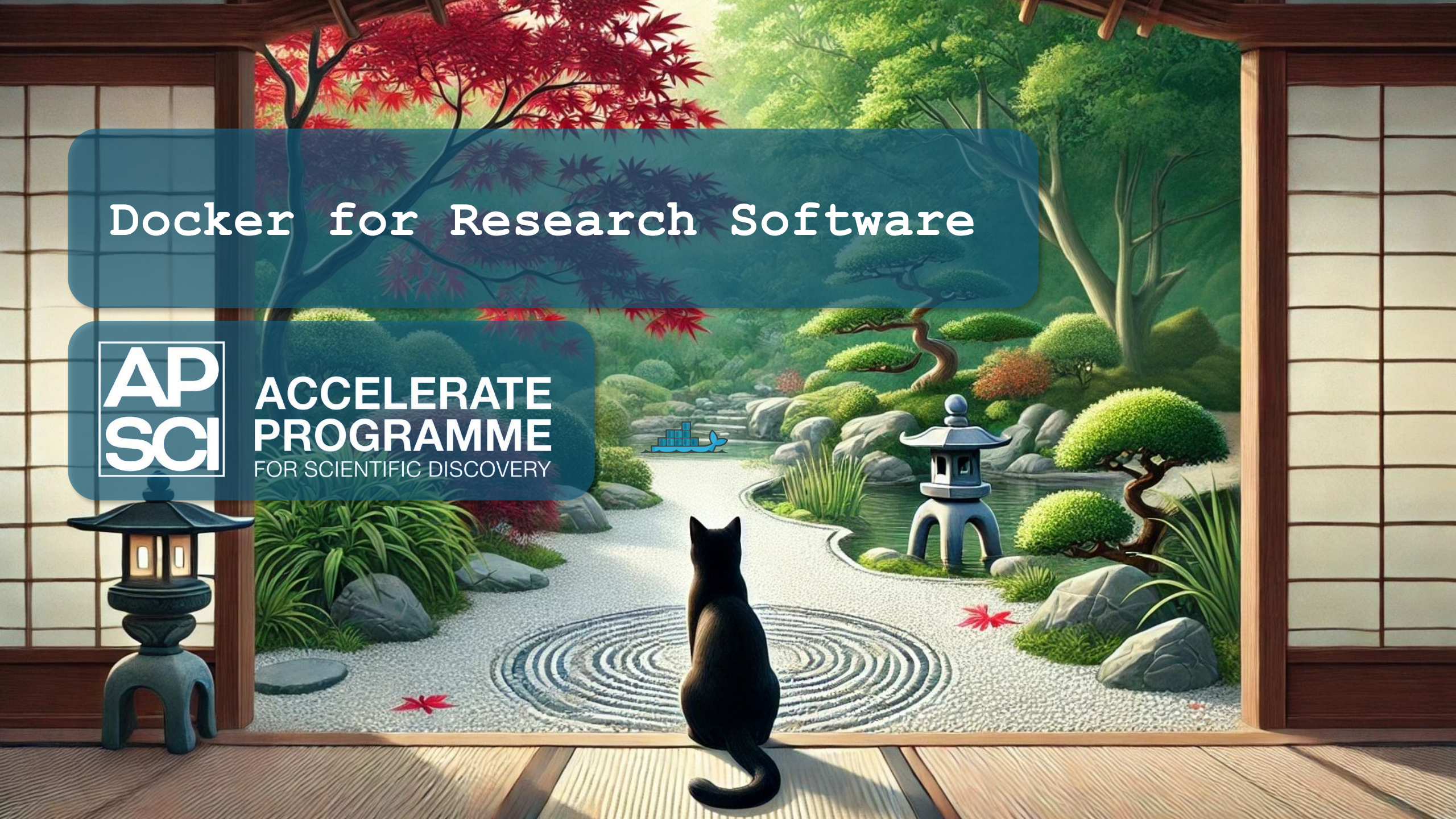
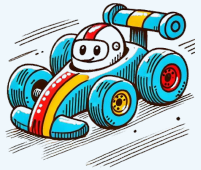

Docker for Research Software
AP SCI
ACCELERATE PROGRAMME
FOR SCIENTIFIC DISCOVERY

About us

MLE @ Accelerate Science – University of Cambridge

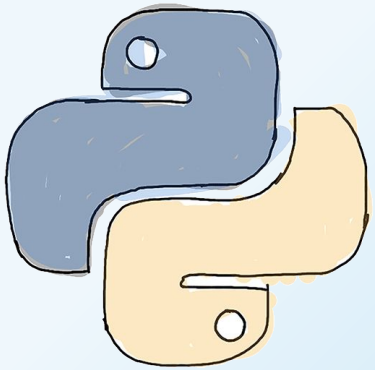
I really like Physics…

…and machine learning…

…and race cars 

Setting the scene

Why I hate python



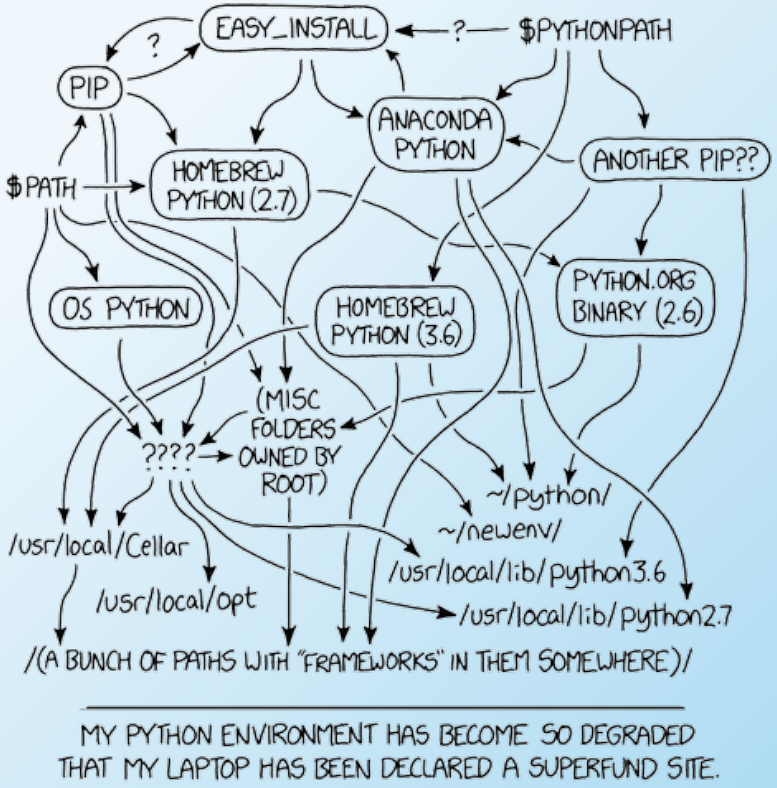
No clear packaging route.

No automatic environments.

So. Many. Tools.

conda

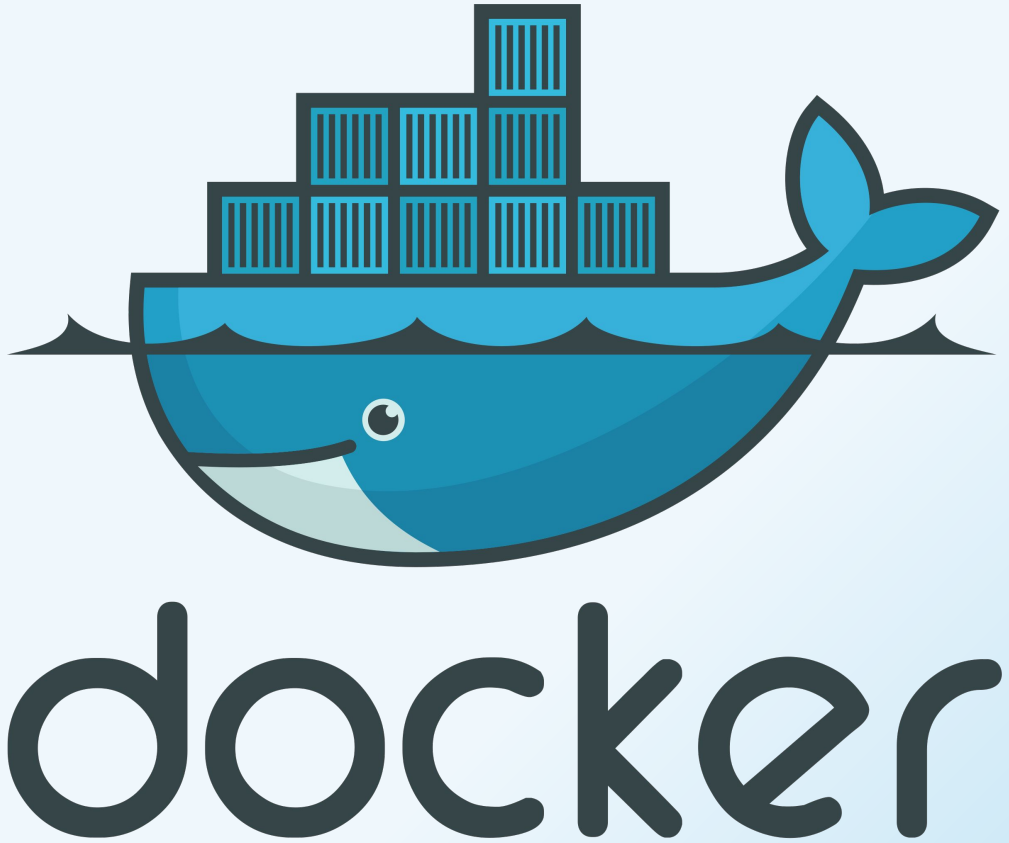
...but it's the best we've got...



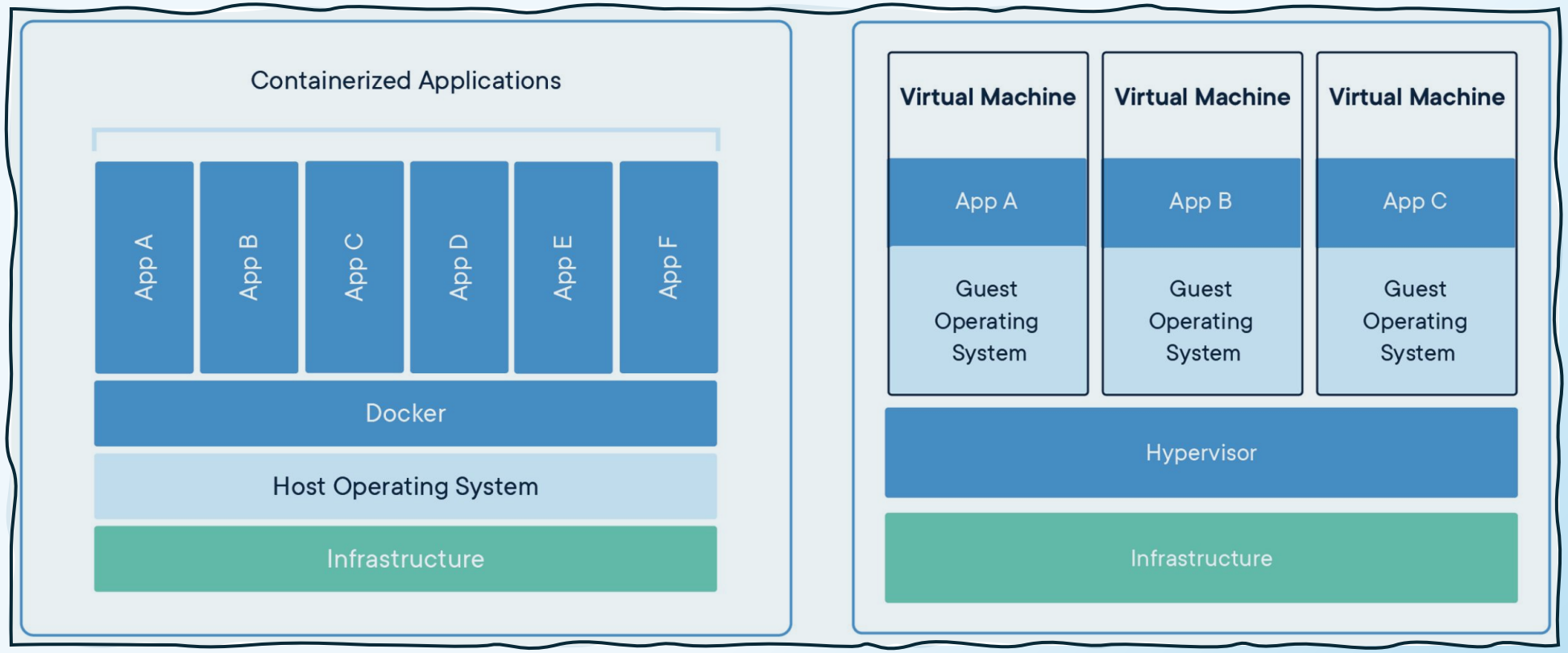
What do we want?

1. Reproducibility
2. Portability
3. Versioning / redundancy
4. Isolation
5. Satisfaction

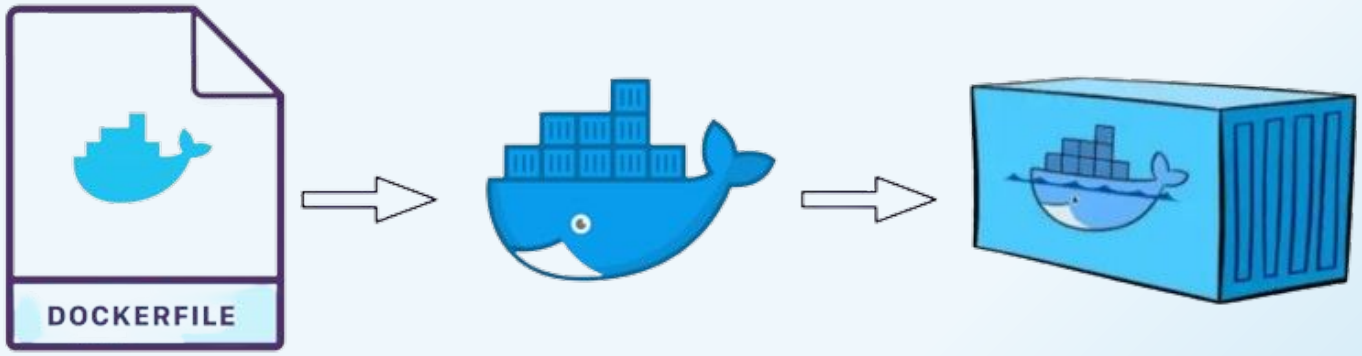
Enter Docker



What is docker?

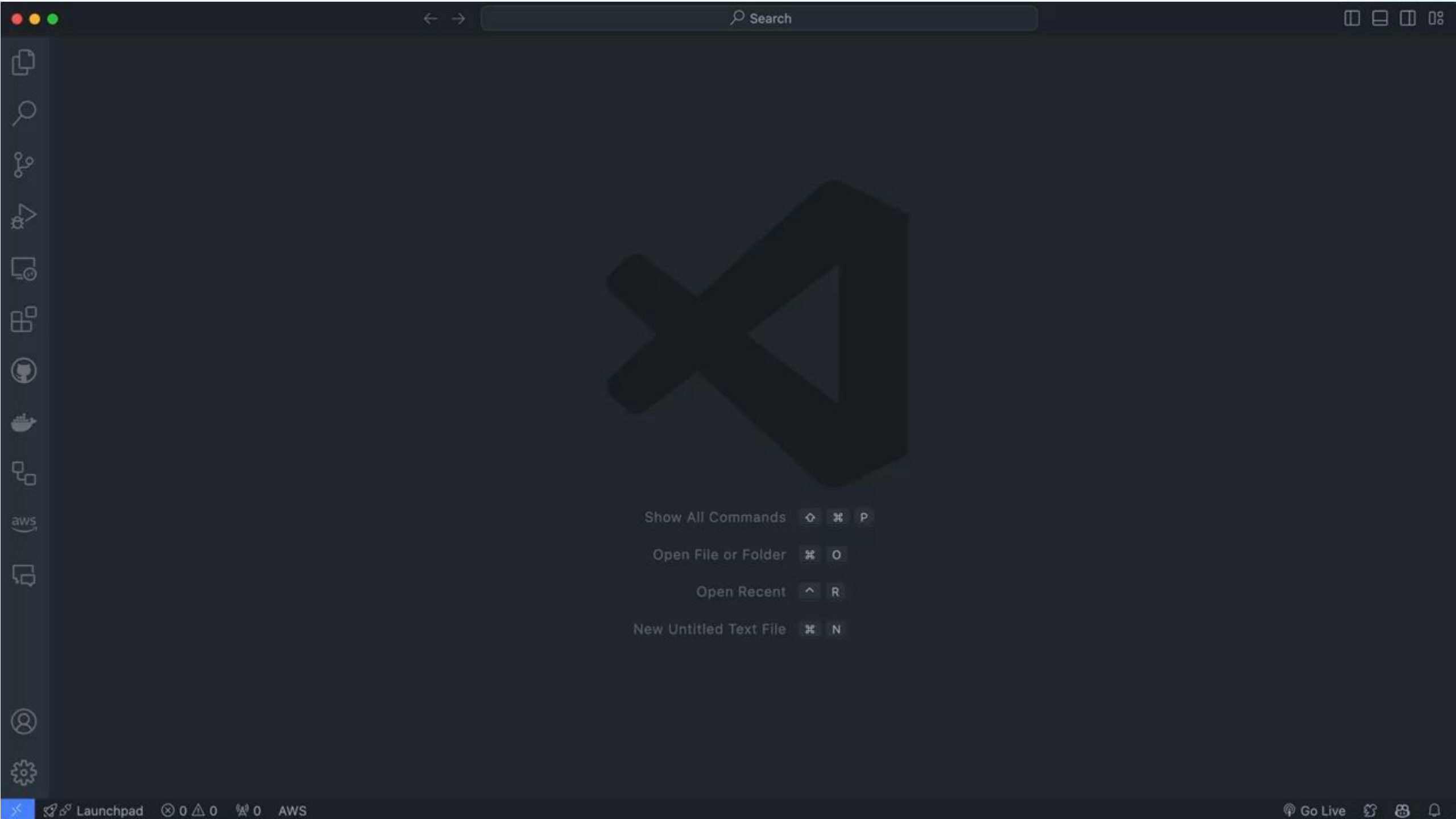


What is docker?



Dockerfile image container

rkd@Ryans-MacBook-Pro-4:~
⌥⌘1
~


Search
Show All Commands ⇧ ⌘ P
Open File or Folder ⌘ O
Open Recent ^ R
New Untitled Text File ⌘ N
Launchpad
0 0
0
AWS
Go Live

Daily driving Docker…

…as a development environment

Docker as a development environment

It’s obviously hard and with no benefits

It’s obviously easy but there are no benefits

Docker is just hard

People don’t know

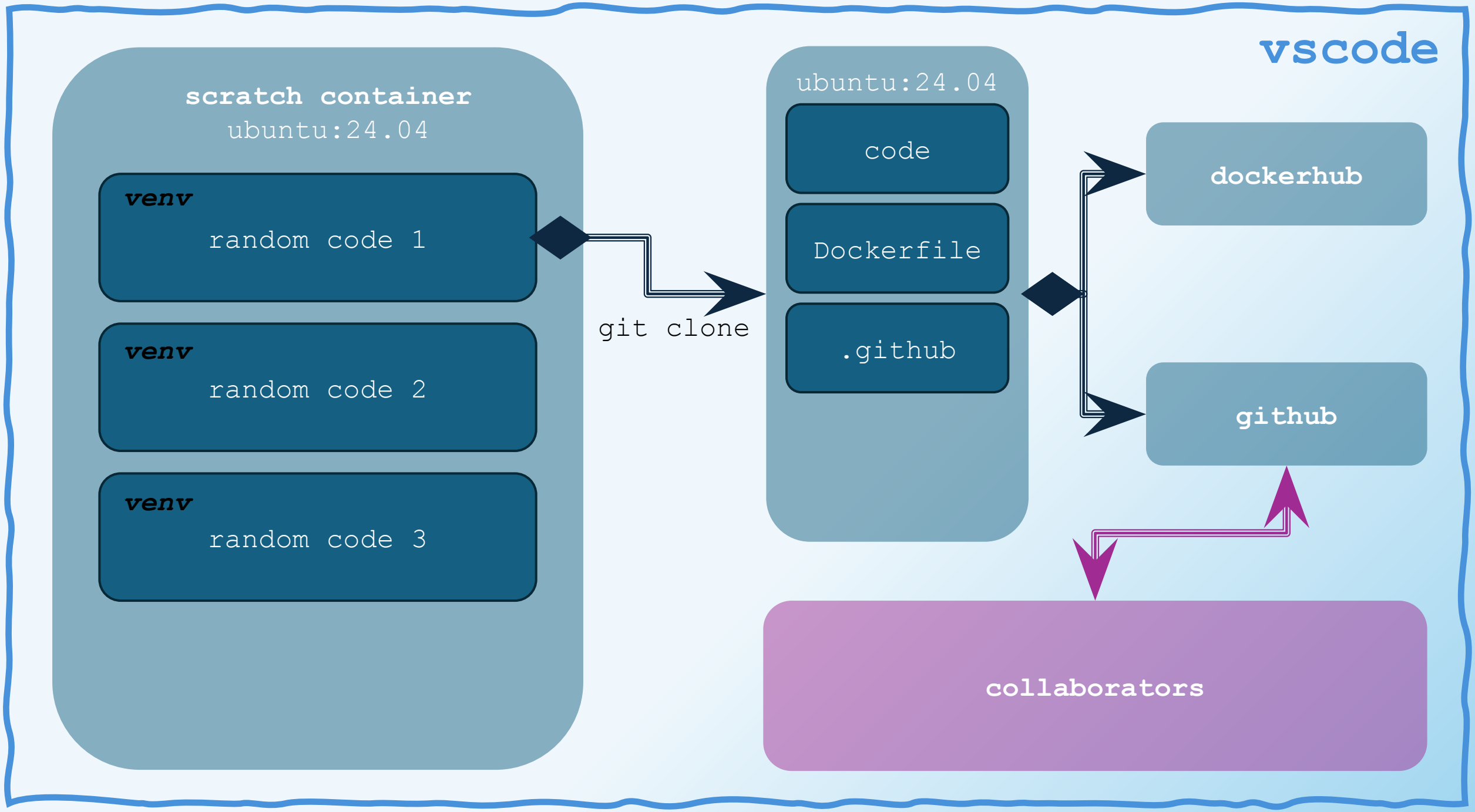
Why?

All the benefits of Docker.

Clean personal environment.

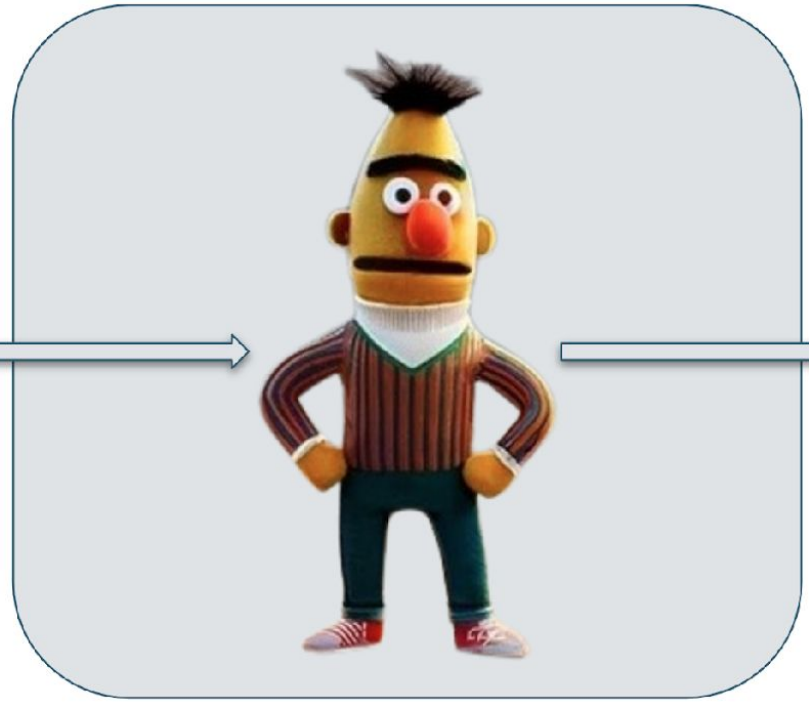
Develop a deeper understanding of Docker.

The Workflow


vscode
scratch container
ubuntu:24.04
venv
random code 1
venv
random code 2
venv
random code 3
git clone
ubuntu:24.04
code
Dockerfile
.github
dockerhub
github
collaborators

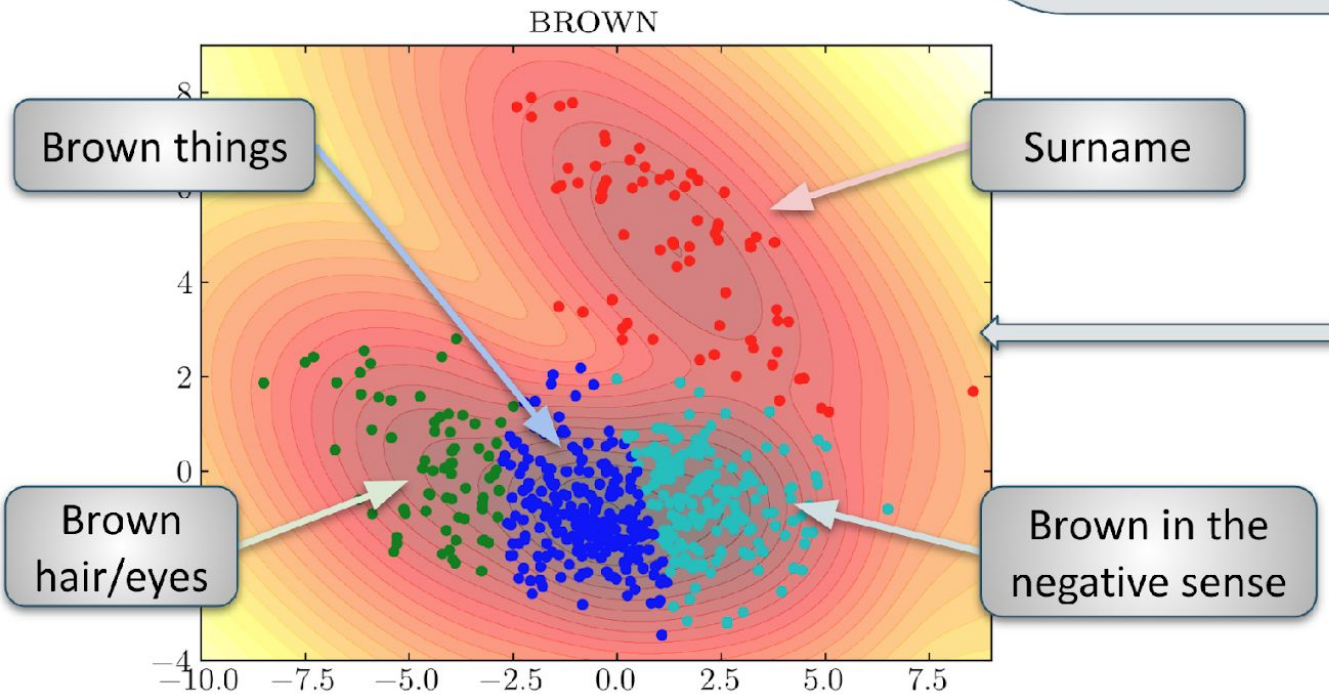
Examples

‘the’	‘[CLS]’	101
‘quick’	‘the’	1996
‘brown’	‘quick’	4248
‘fox’	‘brown’	2829
‘jumps’	‘fox’	4419
	‘jumps’	14523
	‘[SEP]’	102



$X \in \mathbb{R}^{(2C+1)\times d}$

0.17	0.01		0.04
0.33	0.54		0.27
0.71	0.57	$e = X_w \in \mathbb{R}^{1\times d}$	0.49
0.65	0.91	○ ○ ○	0.11
0.31	0.45		0.22
0.43	0.71		0.42
0.96	0.19		0.69

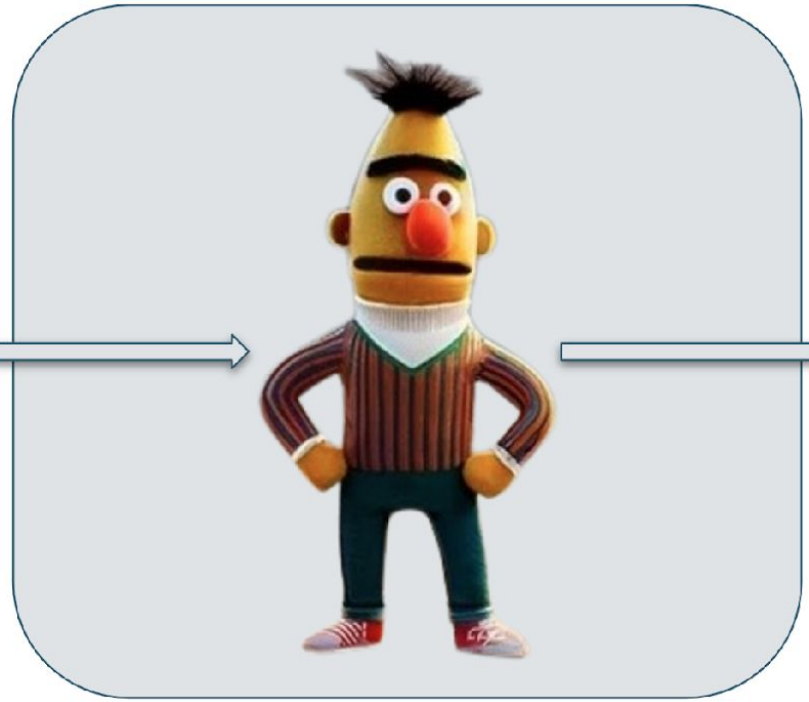


0.32	0.51	○ ○ ○	0.99
0.05	0.13		0.40
0.65	0.91	○ ○ ○	0.11
⋮			⋮
0.65	0.01	○ ○ ○	0.01

e

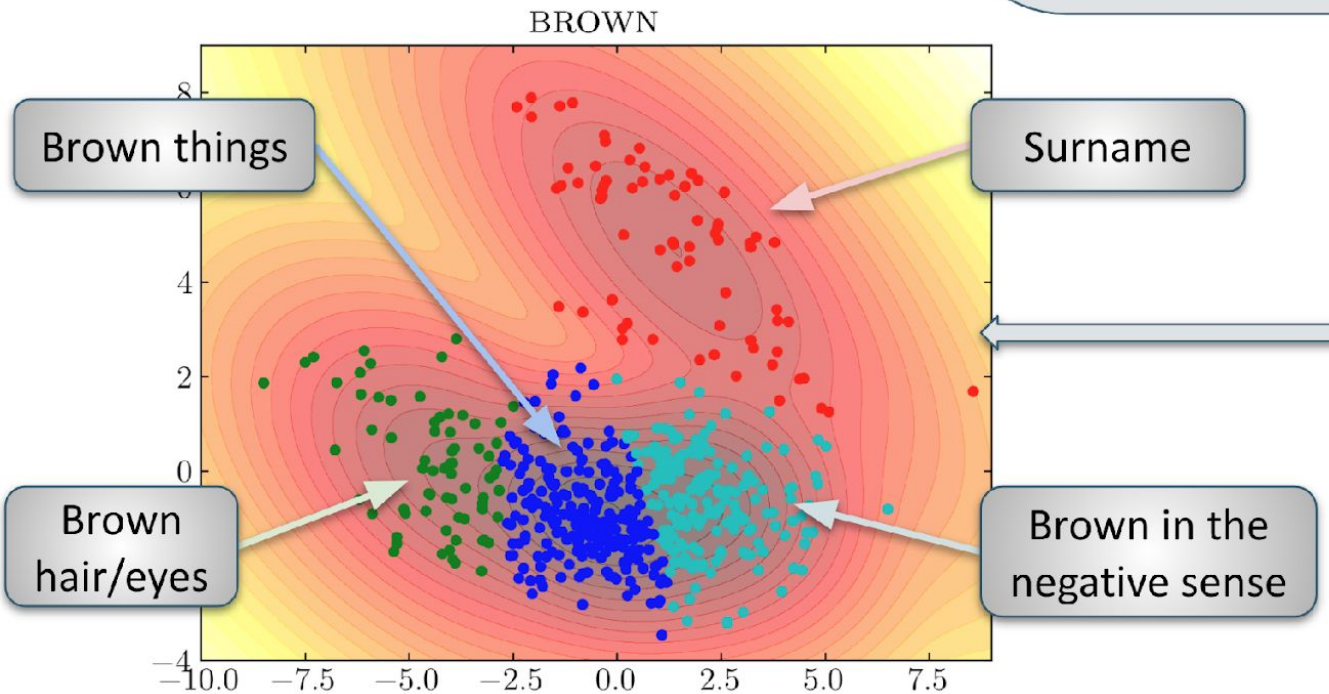
$E \in \mathbb{R}^{n\times d}$

‘the’	‘[CLS]’	101
‘quick’	‘the’	1996
‘brown’	‘quick’	4248
‘fox’	‘brown’	2829
‘jumps’	‘fox’	4419
	‘jumps’	14523
	‘[SEP]’	102



$X \in \mathbb{R}^{(2C+1)\times d}$

0.17	0.01		0.04
0.33	0.54		0.27
0.71	0.57	$e = X_w \in \mathbb{R}^{1\times d}$	0.49
0.65	0.91	○ ○ ○	0.11
0.31	0.45		0.22
0.43	0.71		0.42
0.96	0.19		0.69



0.32	0.51	○ ○ ○	0.99
0.05	0.13		0.40
0.65	0.91	○ ○ ○	0.11
⋮			⋮
0.65	0.01	○ ○ ○	0.01

e

$E \in \mathbb{R}^{n\times d}$

License GPL v3 | issues repo not found | contributors repo not found | issues repo not found | PRs welcome

Star repo not found | Watch repo not found | Fork repo not found | Follow 12 | Follow @AccelerateSci

ACCELERATE PROGRAMME FOR SCIENTIFIC DISCOVERY

REPO TITLE

REPO DESCRIPTION

Start the Course »

Get Basic Code · Report Bug · Request Feature

▶ Table of Contents

(back to top)

How to Run

You can simply copy the git repository and install the dependencies with pip in a virtual environment:

```
pip install -r requirements.txt
```

Or you can use Docker. Pull the image from the Accelerate Science Docker Hub repository:

```
docker pull acceleratescience/website-pages-template
```

To enter the container, run

```
docker run -it -p 8080:8080 website-pages-template /bin/bash
```

I suggest using the VSCode Remote - Containers extension to develop in the container. This will allow you to use the VSCode editor on your local machine while running the code in the container.

To serve locally, run

```
mkdocs serve
```

Course material

Running workshops can be painful

What is your favorite OS...?

...and why is it Linux?

This template enables us to easily develop consistent courses

Problems

Gotchas

Docker for desktop is not great...

...use orbstack instead.

1. Dynamic memory and compute allocation
2. Lighter and faster

Gotchas

Added complexity with GPUs…

…but not really

- Nvidia Container Toolkit
- Most cloud GPU instances come with this:

```
~$ docker run --gpus all my-project
```

Gotchas

Added complexity with remoting into cloud instances…

…but again…not really

- `ssh` into remote
- Pull image and run
- Use vscode dev containers

Gotchas

Different platforms…

…and this is actually sometimes a problem

- amd64 vs arm64
- Need to do long multi-platform builds