How to set up a container in your own computer using DOCKER for reproducible results

Presenting: Esther Camacho

Thursday 23/11/23 at 1:30pm

UNITAT
D'ESTADÍSTICA I
BIOINFORMÀTICA

VHIR
Vall d'Hebron
Barcelona Campus Hospitalari

What is a container?

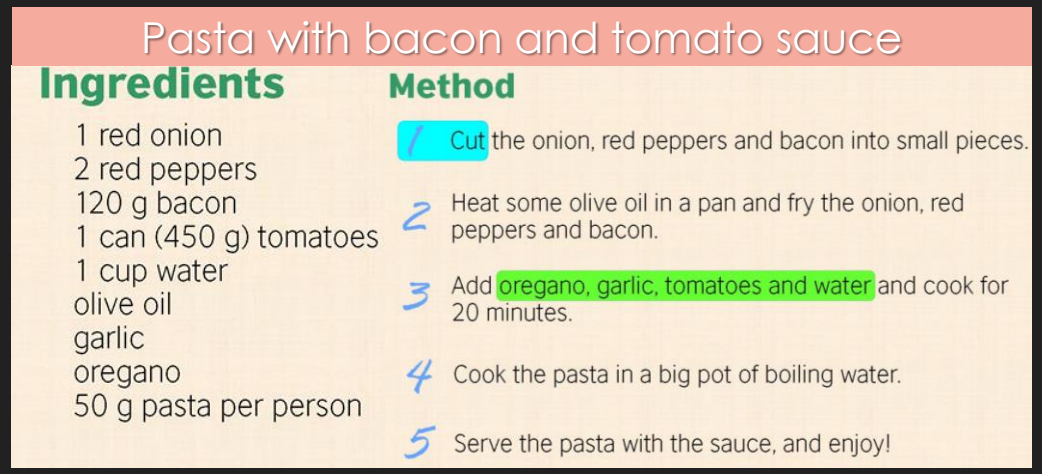
- **Virtualized environment**: an isolated file system accessible from a host computer
- They **only carry the essential information** to perform their task (specific purposes)

Advantages of working with containers

- **Reproducibility**: We avoid problems of dependencies
- **Lightweight**
- We can **share the environment**:
  - Multiple containers at the same time
  - Containers can communicate among them

Key concepts

IMAGE



Pasta with bacon and tomato sauce

Ingredients

1 red onion
2 red peppers
120 g bacon
1 can (450 g) tomatoes
1 cup water
olive oil
garlic
oregano
50 g pasta per person

Method

1. Cut the onion, red peppers and bacon into small pieces.
2. Heat some olive oil in a pan and fry the onion, red peppers and bacon.
3. Add oregano, garlic, tomatoes and water and cook for 20 minutes.
4. Cook the pasta in a big pot of boiling water.
5. Serve the pasta with the sauce, and enjoy!

- Read-only
- Stored on longer term
- Can be used as a base

CONTAINER

- Based on the image
- Short-lived
- Usually only minor adjustments

Docker

- Most popular container software
- GUI on Linux, MacOS and Windows
- Great for container development
- Very large repository (**docker hub**) with base images such as:
  - ubuntu
  - r-base/rstudio server
  - python
  - conda
  - etc.

Other container software

- Singularity (HPC)
- Shifter (HPC)
- Charliecloud (HPC)
- Podman (daemonless docker)

podman

Differences between containers and virtual machines

Containers

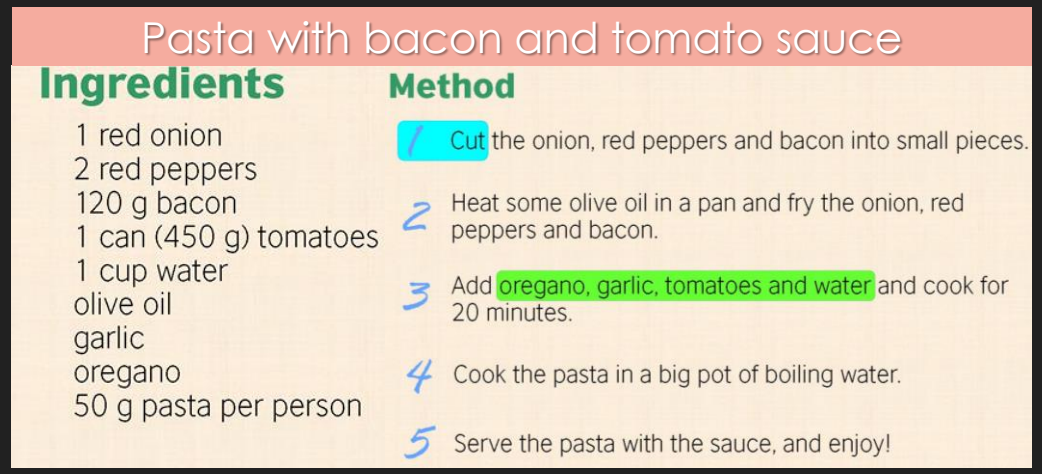
- Share host system's OS kernel
- Better resource utilization and efficiency
- Lower isolation, but still isolated at user space level
- Lower overhead, faster startup times
- Portable, but dependent on containerization

Virtual machines

- Emulate entire hardware environment
- Higher resource utilization inefficiencies
- Strong isolation between virtual machines
- Performance overhead due to emulation
- Highly portable

Managing containers

IMAGE



Pasta with bacon and tomato sauce

Ingredients

1 red onion
2 red peppers
120 g bacon
1 can (450 g) tomatoes
1 cup water
olive oil
garlic
oregano
50 g pasta per person

Method

1. Cut the onion, red peppers and bacon into small pieces.
2. Heat some olive oil in a pan and fry the onion, red peppers and bacon.
3. Add oregano, garlic, tomatoes and water and cook for 20 minutes.
4. Cook the pasta in a big pot of boiling water.
5. Serve the pasta with the sauce, and enjoy!

- Read-only
- Stored on longer term
- Can be used as a base

CONTAINER

- Based on the image
- Short-lived
- Usually only minor adjustments

Managing containers

IMAGE

Pasta with bacon and tomato sauce

Ingredients

1 red onion
2 red peppers
120 g bacon
1 can (450 g) tomatoes
1 cup water
olive oil
garlic
oregano
50 g pasta per person

Method

1. Cut the onion, red peppers and bacon into small pieces.
2. Heat some olive oil in a pan and fry the onion, red peppers and bacon.
3. Add oregano, garlic, tomatoes and water and cook for 20 minutes.
4. Cook the pasta in a big pot of boiling water.
5. Serve the pasta with the sauce, and enjoy!

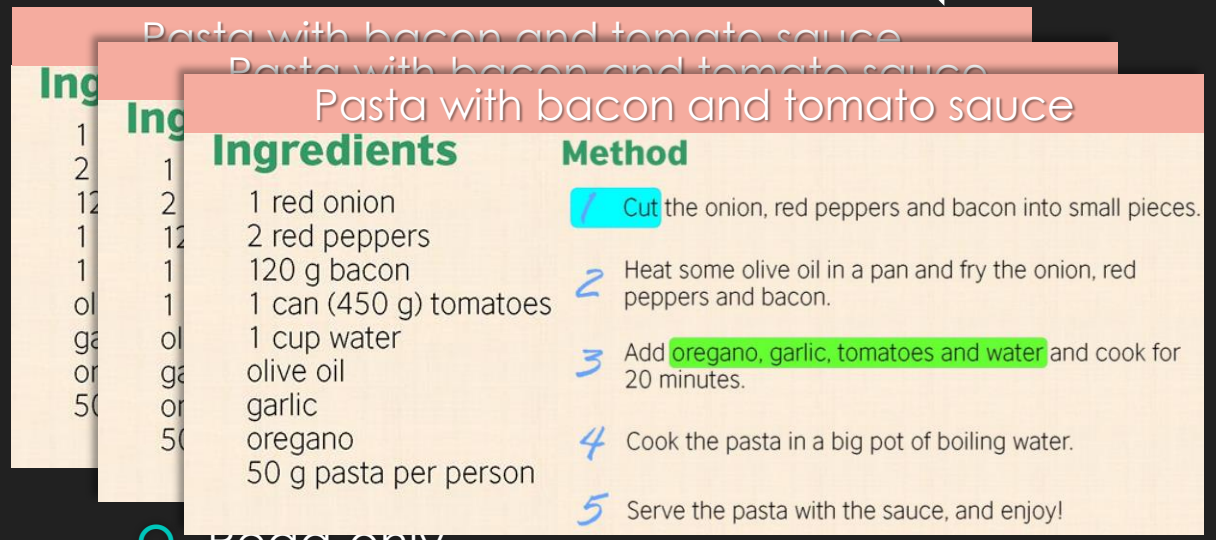
- Read-only
- Stored on longer term
- Can be used as a base

CONTAINER

- Based on the image
- Short-lived
- Usually only minor adjustments

Managing containers

IMAGE

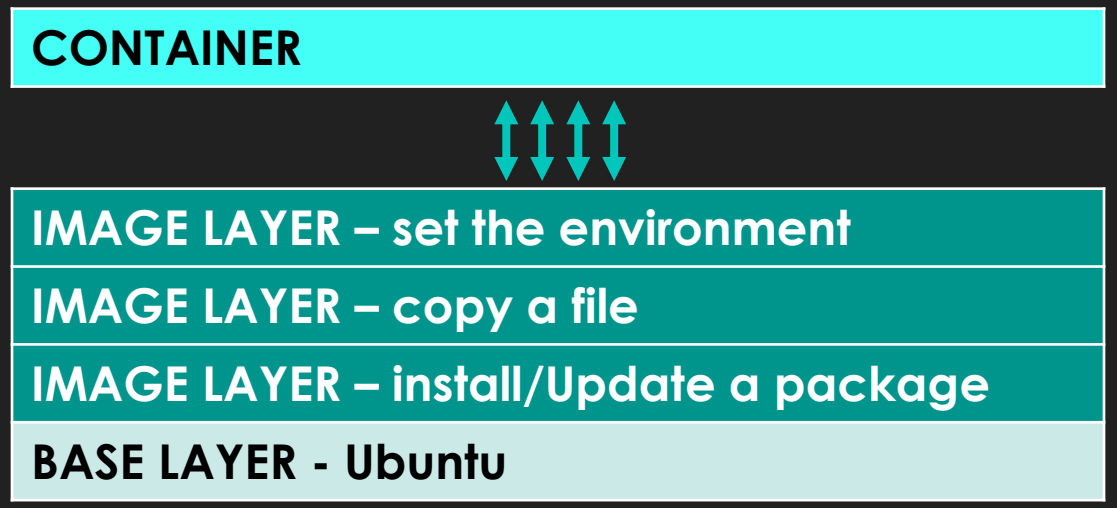


- Read-only
- Stored on longer term
- Can be used as a base

CONTAINER

- Based on the image
- Short-lived
- Usually only minor adjustments

Layers


CONTAINER
IMAGE LAYER – set the environment
IMAGE LAYER – copy a file
IMAGE LAYER – install/Update a package
BASE LAYER - Ubuntu
Container: writeable
Image: read-only

Creating an image

- From a **Dockerfile**
- From a **container**: with *docker commit*

Sharing an image

- **docker hub** (open to the world)
  - Command: ***docker push*** (upload), ***docker pull*** (download)
- command *docker save*
- Dockerfile (source code)

Frequently used features

- **Mounting directories**
  - **Bind-mount** Make a directory on the host available to the container
  - **Volume** Disk space reserved and managed by docker
- **Managing identities**
- **Mapping ports**
  - **Display browser content** (eg. Rstudio server or any other web server)
  - **Published at [IP]:[PORT]** (eg. 127.0.0.1:8000)
  - **Forward the port from the container to port on the host:** docker run –p 80:8000

publish port 8000 in the container at port 80 on the host

Creating a container with docker…

Demo 1

Working with Dockerfiles

Demo 2

Create a Dockerfile

Demo 2.1

Create an image from a Dockerfile

Demo 2.2

Push/pull an image to docker hub

Demo 2.3

Run a container

Demo 2.4

References

- The material of this course is based on the public SIB material of 'Introduction to containers' : https://sib-swiss.github.io/containers-introduction-training/2023.4/course_material/introduction_containers/
- Link to dockerhub: https://hub.docker.com/
- Images of example for cooking recipe: https://learnenglishteens.britishcouncil.org/skills/writing/a2-writing/recipe