Docker Container

Rowel Atienza, PhD

University of the Philippines

github.com/roatienza

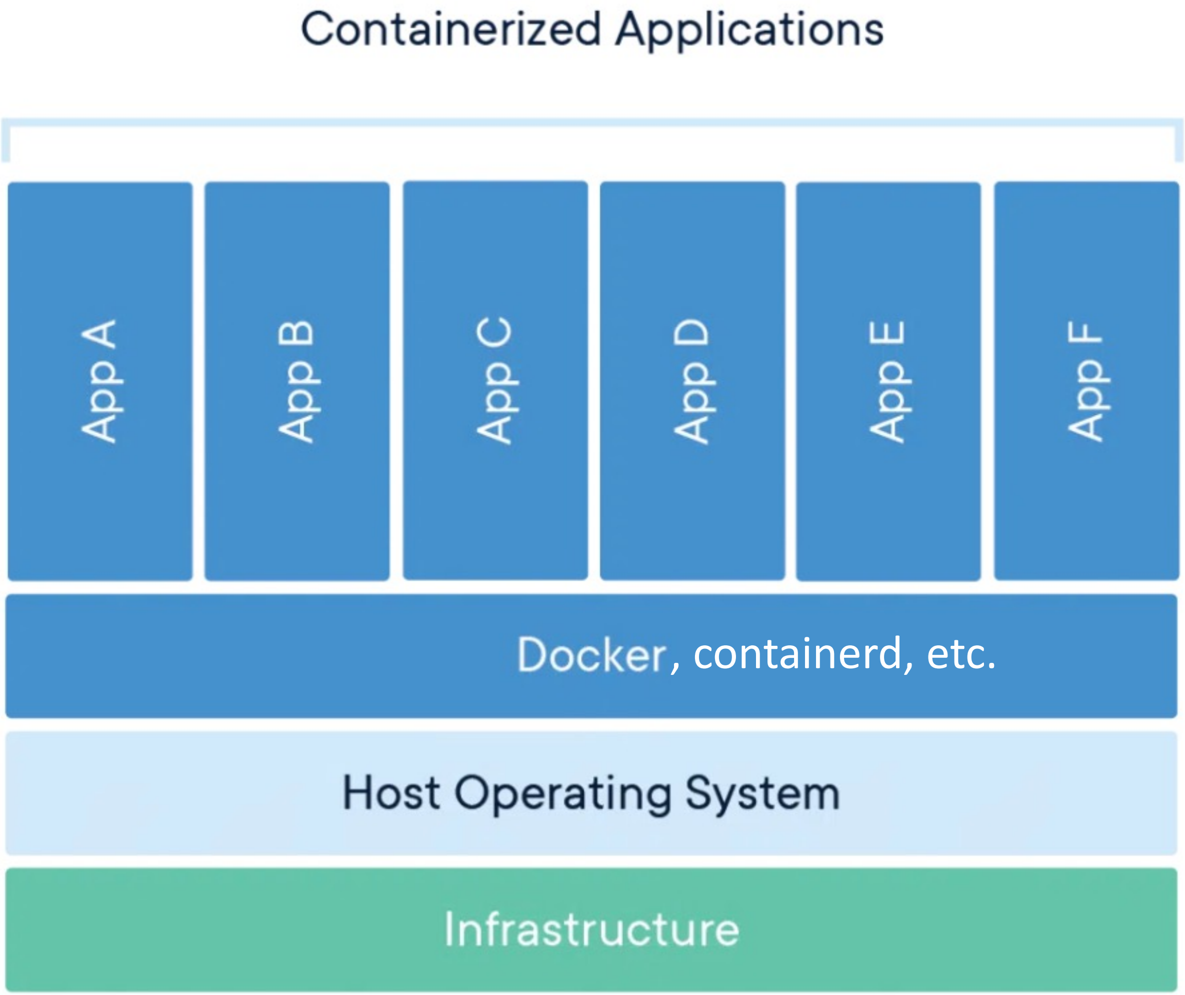
2023

Why Containers and Why Docker

- Container - Packaged code and all its dependencies so applications runs quickly and reliably from one computing environment to another
- Docker - lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings

docker.com

Container and Docker



Containerized Applications

App A

App B

App C

App D

App E

App F

Docker, containerd, etc.

Host Operating System

Infrastructure

ML/AI Model Serving & Apps

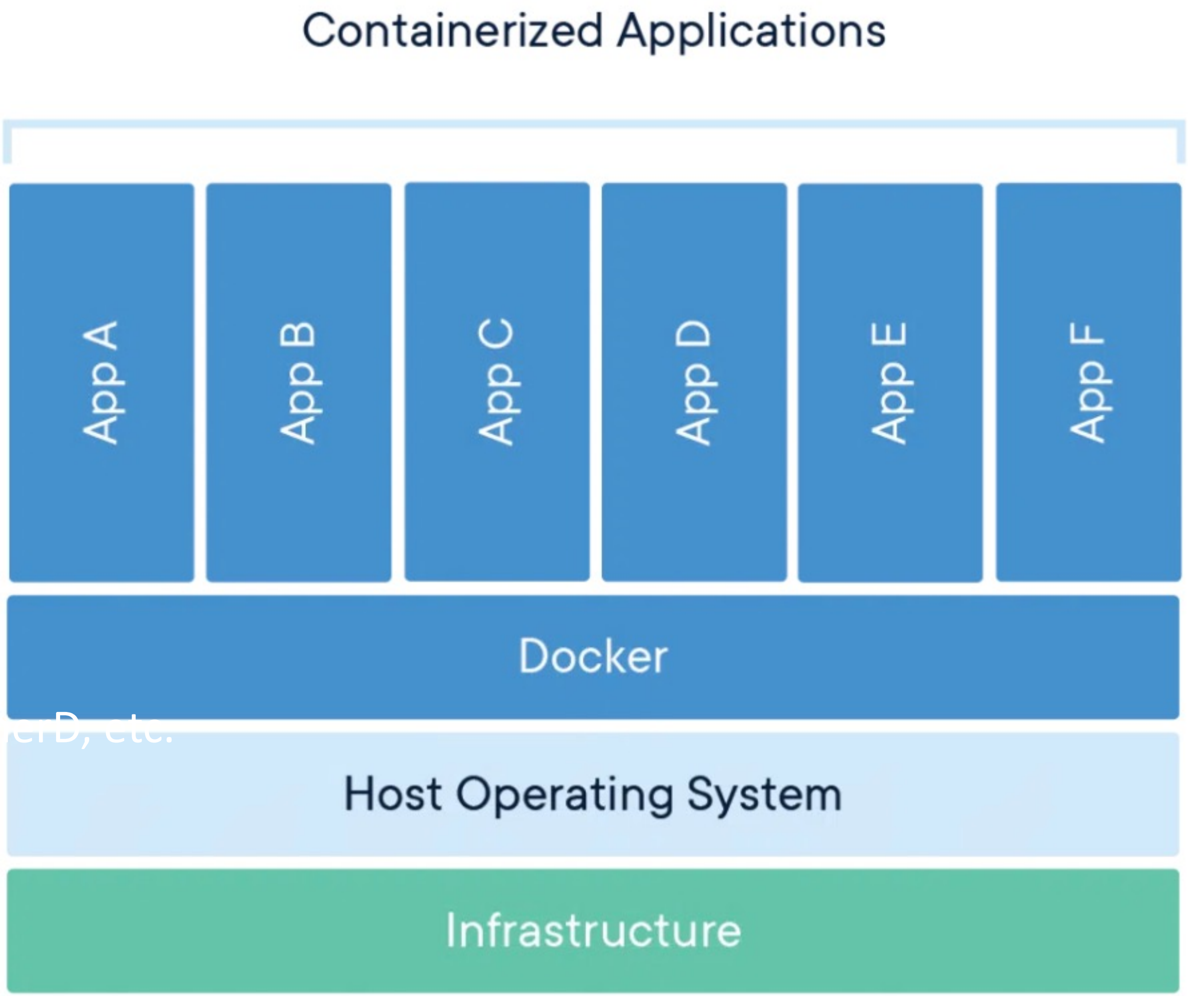
Engine

Linux or Windows

Cloud or On-Premise

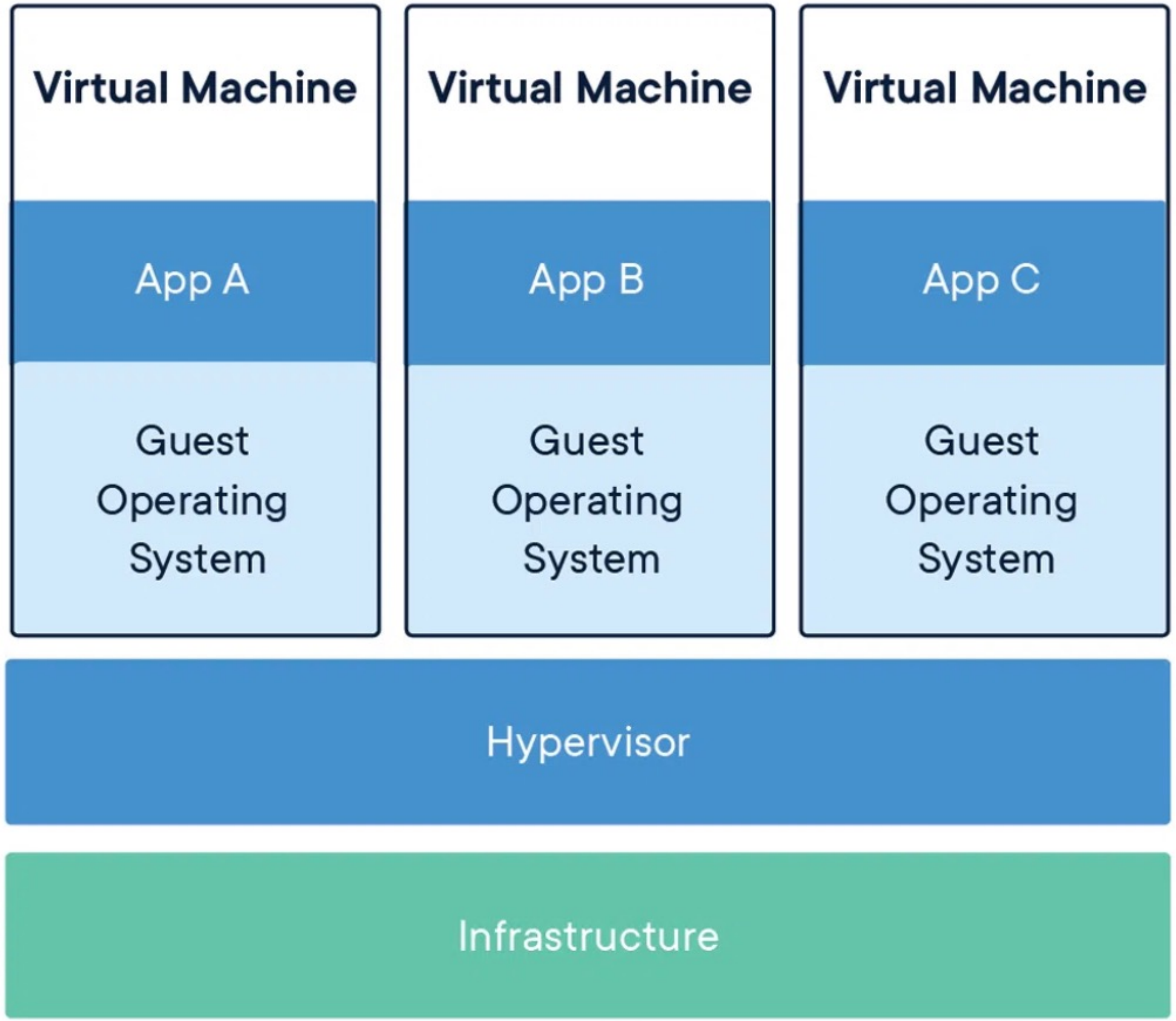
Container

- Virtualizes the OS
- Small footprint (Dependencies footprint). Efficient
- Maximizes OS use
- Fast OS boot-up and container start-up
- Target: Applications
- e.g. Docker, ContainerD



Virtualization

- Virtualizes the hardware
- Big footprint (OS footprint)
- Maximizes hardware use
- Slow OS boot-up and Guest OS start-up
- Target: Full-server deployment
- e.g. Xen, Parallels



Container + Virtualization

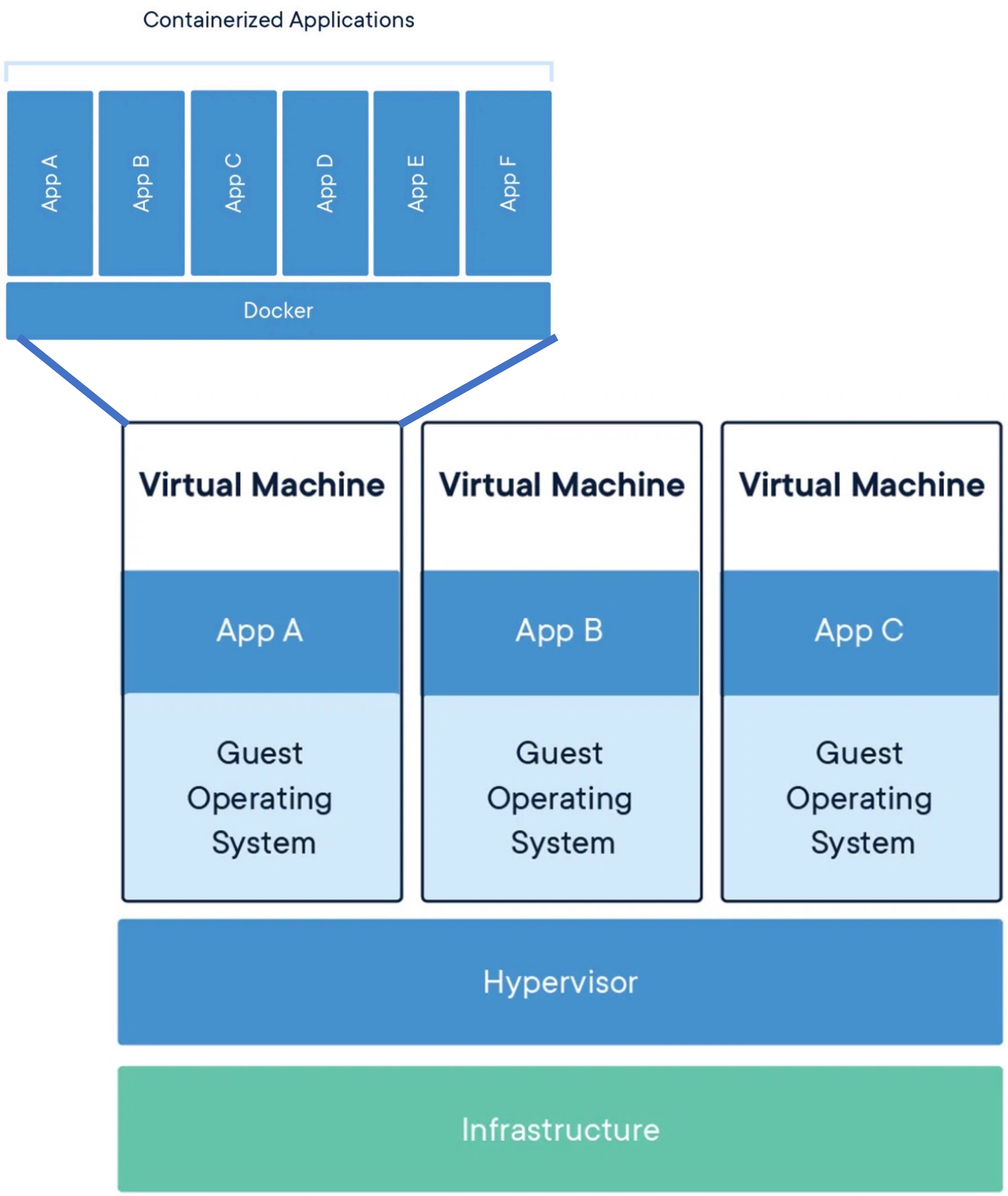
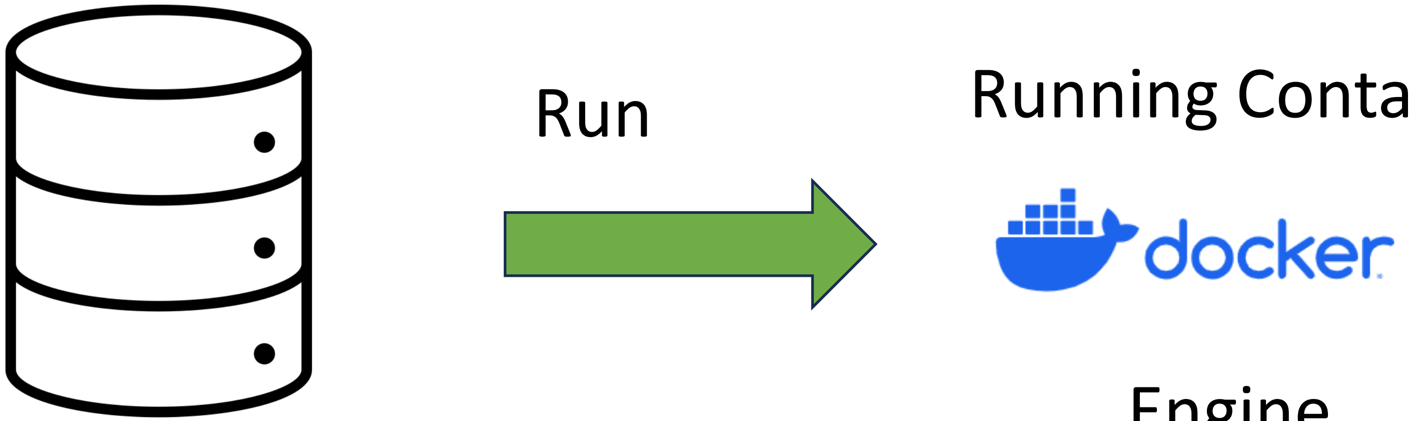


Image to Container



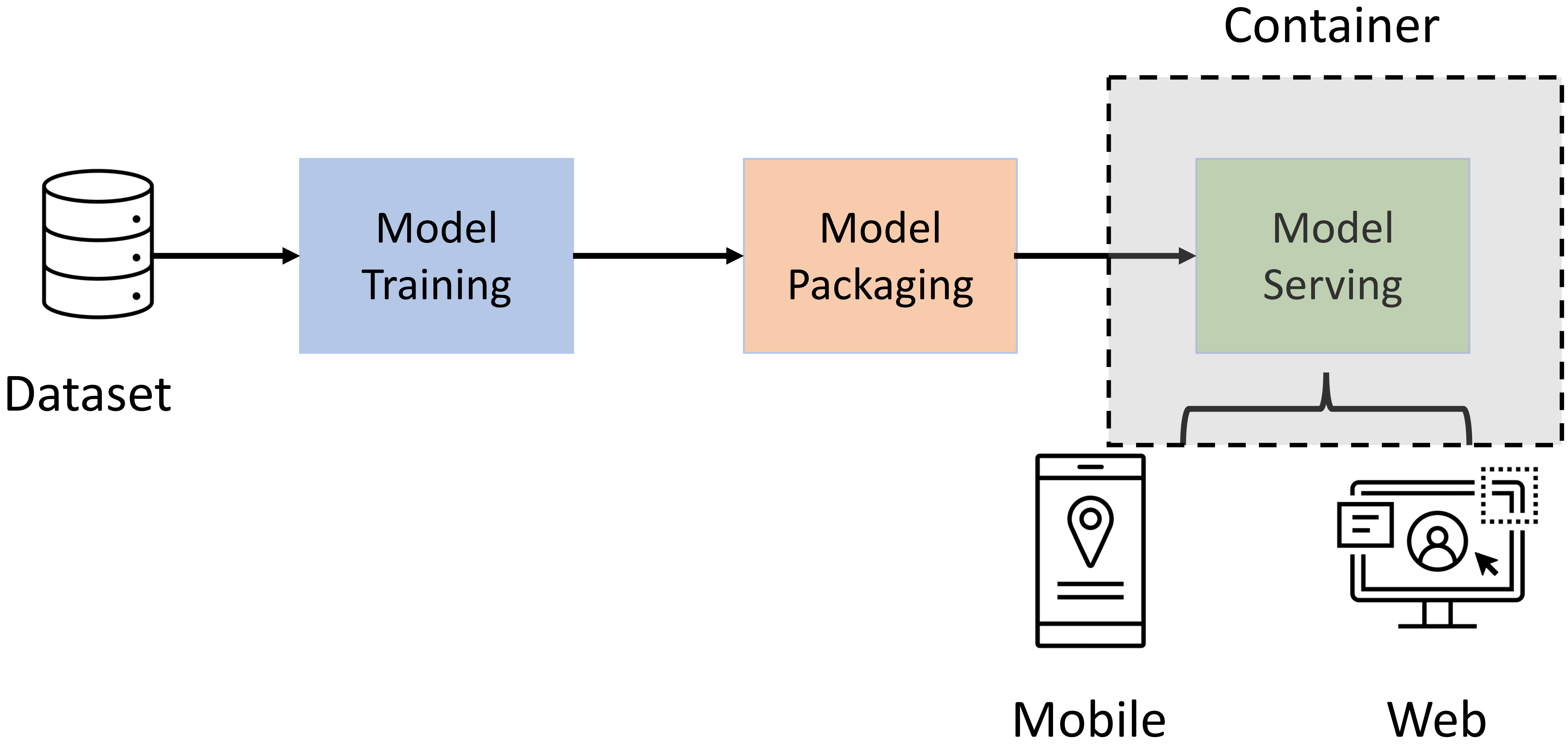
Run

Running Container

Engine

Image in Filesystem

Model Deployment Pipeline



Gradio or Streamlit Front-end

Container Standards

- *containerd* - industry-standard container runtime

NVIDIA Containers

NVIDIA NGC

- NGC – NVIDIA’s ecosystem for software, enterprise services, tools, etc
  - Target Deployment - Cloud, On-premise (DGX), Device (Jetson)
  - Containers - SDK Containers for AI and data science software, tuned, tested, and optimized by NVIDIA
  - Collections - Ready to deploy AI applications – LLMs, Vision, Speech, Recommendation System, etc
  - Pre-trained Models – Ready to use

NGC Docker Containers

- Docker is used for creating, deploying, and running containers

docker

NVIDIA NGC

Using Docker on NGC for Jetson

Downloading the Container

```
docker pull [OPTIONS] NAME[:TAG|@DIGEST]
```

```
e.g.
docker pull dustynv/llama_cpp:ggml-r35.4.1
```

JetPack Version

```
cat /etc/nv_tegra_release
# R35 (release), REVISION: 4.1
```

Therefore, `JetPack ver: 35.4.1`

Adding username to docker group

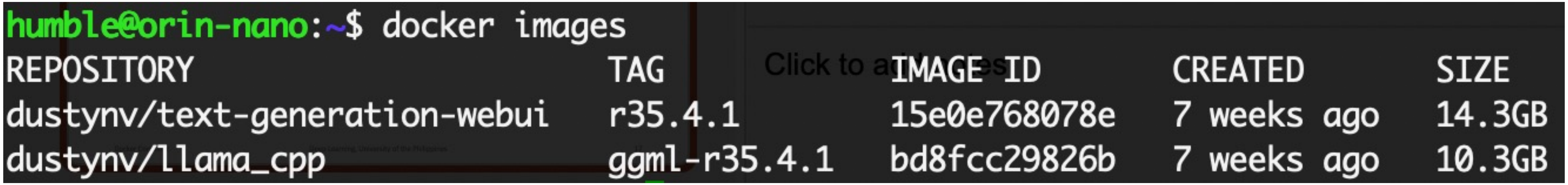
```
sudo usermod -aG docker <username>
```

Note:

You have to re-login for this change to take effect

Docker Images available on your device

```
docker images
```



```
humble@orin-nano:~$ docker images
REPOSITORY                    TAG              IMAGE ID       CREATED       SIZE
dustynv/text-generation-webui r35.4.1          15e0e768078e   7 weeks ago   14.3GB
dustynv/llama_cpp             ggml-r35.4.1     bd8fcc29826b   7 weeks ago   10.3GB
```

Running the Docker image

```
docker run --runtime nvidia -it –rm \
 --network=host CONTAINER:TAG


e.g.
docker run -it --rm --net=host --runtime nvidia \
    -e DISPLAY=$DISPLAY \
    -v /tmp/.X11-unix/:/tmp/.X11-unix \
    dustynv/llama_cpp:ggml-r35.4.1
```

Test `llama_cpp`

Prompt:

```
  The Philippines Boracay island is
```

On Jetson Orin:

```
/opt/llama-cpp-python/vendor/llama.cpp/build/bin# ./main –model
$(huggingface-downloader TheBloke/Llama-2-7B-GGML/llama-2-
7b.ggmlv3.q4_0.bin) --prompt "The Philippines Boracay Island is"  --n-
predict 128 --ctx-size 192 --batch-size 192 --n-gpu-layers 999 --threads
$(nproc)
```

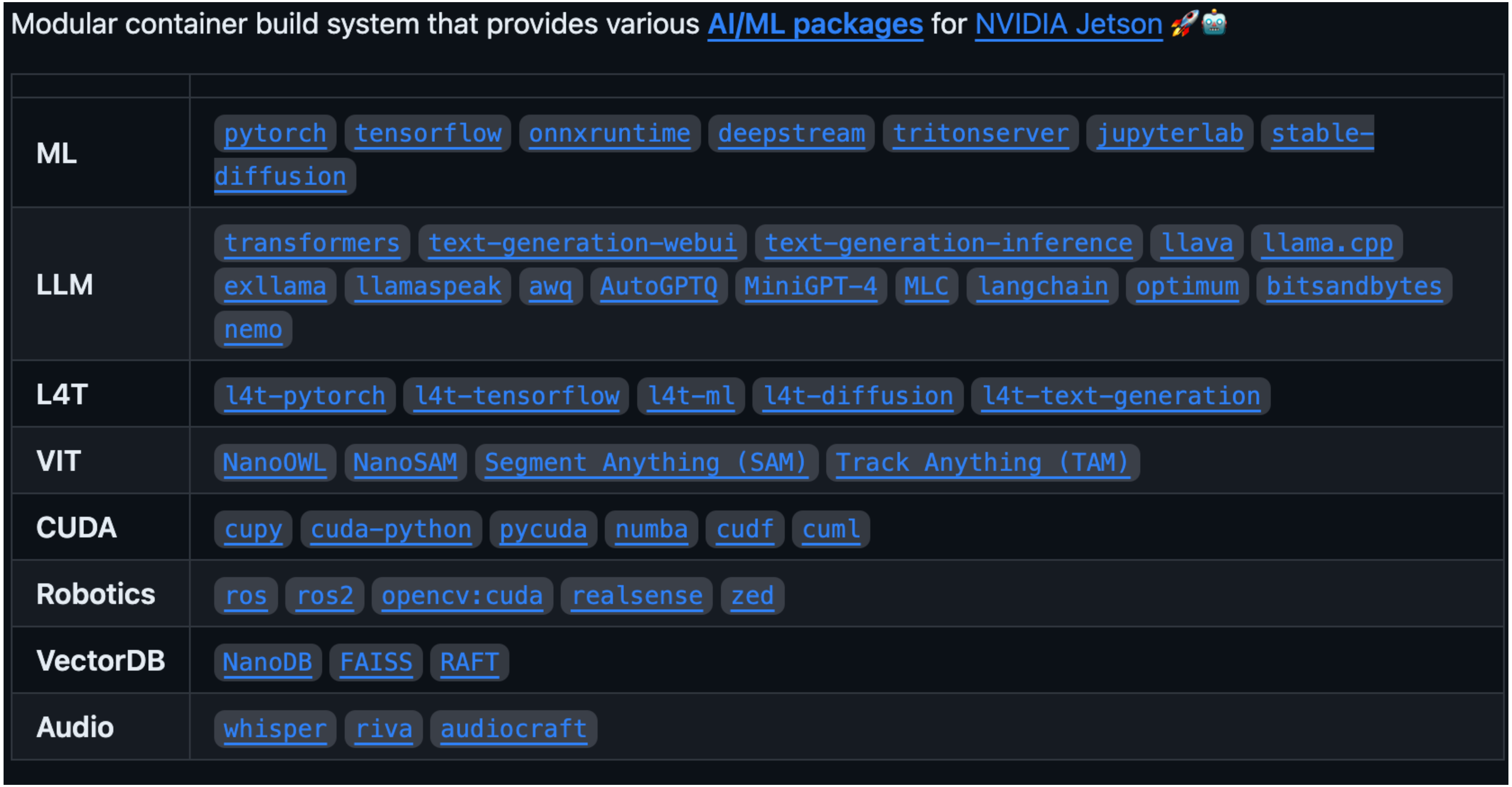
llama_cpp sample response

```
<s> The Philippines Boracay Island is a tropical
paradise and a world-class tourist destination.
Einzelnen Angeboten auf diesem Internetportal ist
die Abrechnung in EUR, CHF oder USD möglich.

Rabindranath Tagore (, ; Bengali: ), known as
Ravīndrānta Thakur in Bengalese and by the sobriquet
Gurudev, was a Bengali polymath who reshaped his
region's culture, politics and religion throughout
his prolific life, and is credited with raising
social consciousness in South Asia.

Aug 23,  ·
```

Machine Learning Containers for Jetson and JetPack

```
sudo apt-get update
sudo apt-get install git python3-pip
git clone --depth=1 \
https://github.com/dusty-nv/jetson-containers
cd jetson-containers
pip3 install -r requirements.txt
```

Packages available



Modular container build system that provides various AI/ML packages for NVIDIA Jetson 🚀🤖

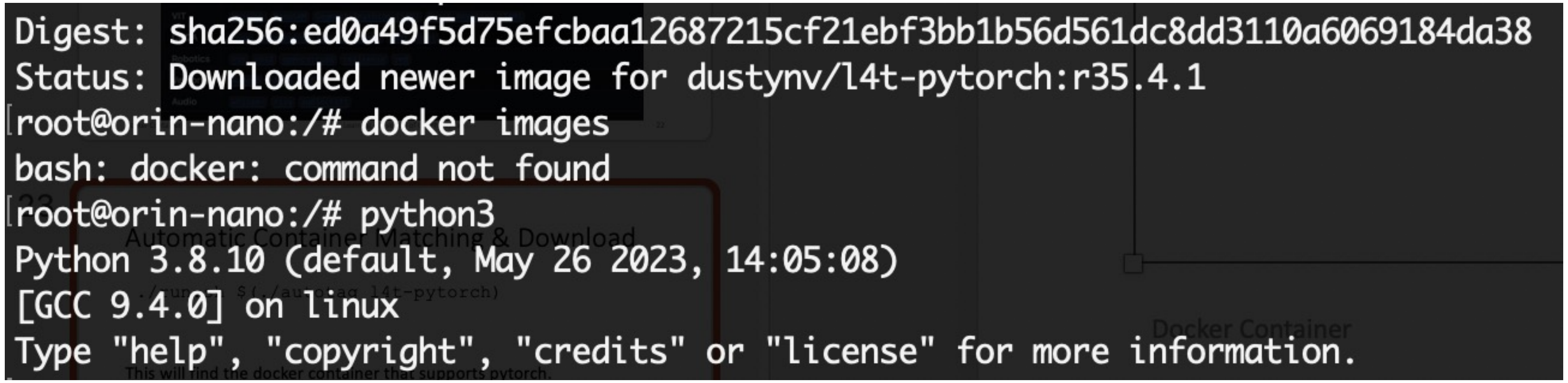
ML	pytorch tensorflow onnxruntime deepstream tritonserver jupyterlab stable-diffusion
LLM	transformers text-generation-webui text-generation-inference llava llama.cpp exllama llamaspeak awq AutoGPTQ MiniGPT-4 MLC langchain optimum bitsandbytes nemo
L4T	l4t-pytorch l4t-tensorflow l4t-ml l4t-diffusion l4t-text-generation
VIT	NanoOWL NanoSAM Segment Anything (SAM) Track Anything (TAM)
CUDA	cupy cuda-python pycuda numba cudf cuml
Robotics	ros ros2 opencv:cuda realsense zed
VectorDB	NanoDB FAISS RAFT
Audio	whisper riva audiocraft

Automatic Container Matching & Download

```
./run.sh $(./autotag l4t-pytorch)
```

This will find the docker container that supports pytorch.

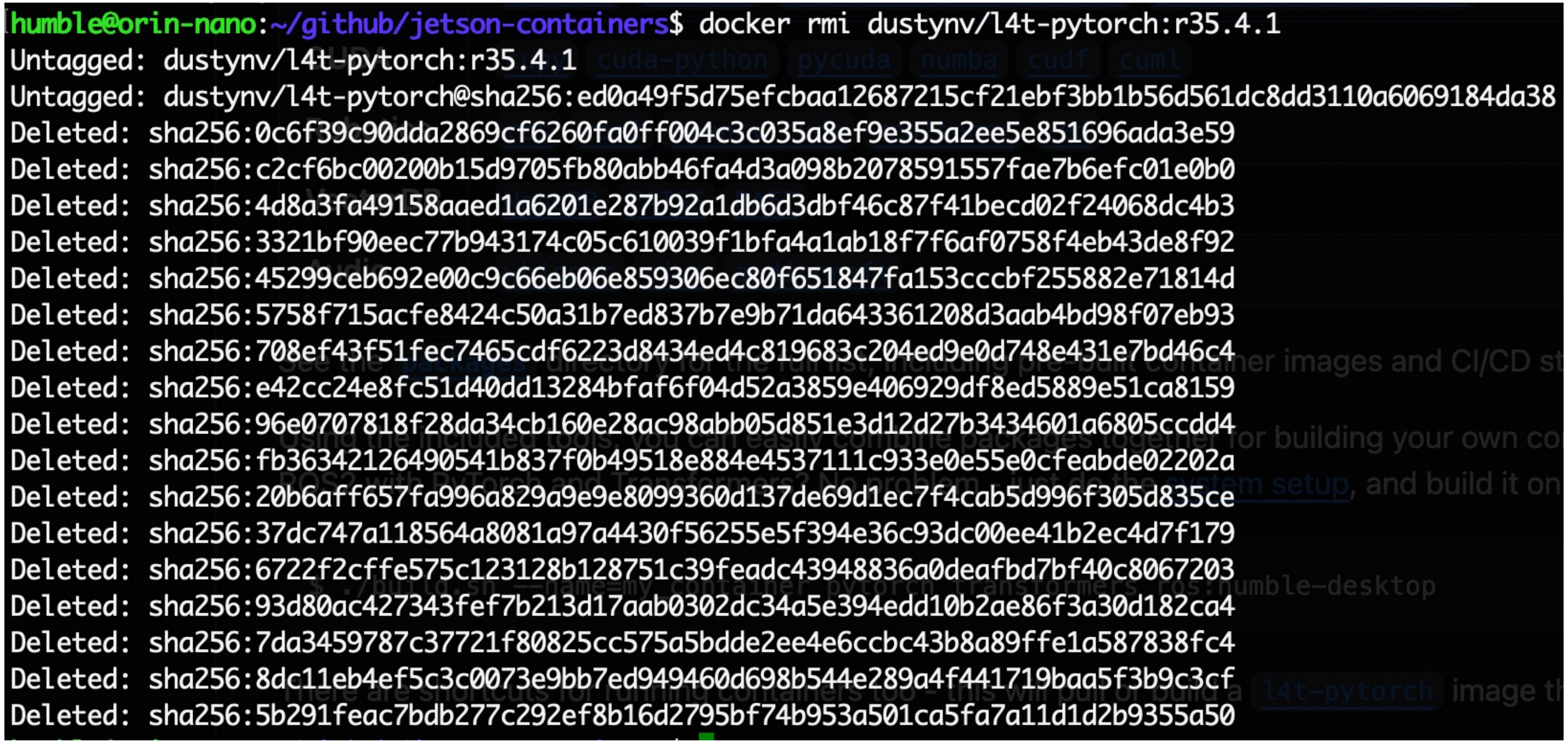
`docker pull` and `run` it if found.



```
Digest: sha256:ed0a49f5d75efcbaa12687215cf21ebf3bb1b56d561dc8dd3110a6069184da38
Status: Downloaded newer image for dustynv/l4t-pytorch:r35.4.1
root@orin-nano:/# docker images
bash: docker: command not found
root@orin-nano:/# python3
Python 3.8.10 (default, May 26 2023, 14:05:08)
[GCC 9.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
```

Removing a docker image

```
docker rmi CONTAINER:TAG
```



```
humble@orin-nano:~/github/jetson-containers$ docker rmi dustynv/l4t-pytorch:r35.4.1
Untagged: dustynv/l4t-pytorch:r35.4.1
Untagged: dustynv/l4t-pytorch@sha256:ed0a49f5d75efcbaa12687215cf21ebf3bb1b56d561dc8dd3110a6069184da38
Deleted: sha256:0c6f39c90dda2869cf6260fa0ff004c3c035a8ef9e355a2ee5e851696ada3e59
Deleted: sha256:c2cf6bc00200b15d9705fb80abb46fa4d3a098b2078591557fae7b6efc01e0b0
Deleted: sha256:4d8a3fa49158aaed1a6201e287b92a1db6d3dbf46c87f41becd02f24068dc4b3
Deleted: sha256:3321bf90eec77b943174c05c610039f1bfa4a1ab18f7f6af0758f4eb43de8f92
Deleted: sha256:45299ceb692e00c9c66eb06e859306ec80f651847fa153cccbf255882e71814d
Deleted: sha256:5758f715acfe8424c50a31b7ed837b7e9b71da643361208d3aab4bd98f07eb93
Deleted: sha256:708ef43f51fec7465cdf6223d8434ed4c819683c204ed9e0d748e431e7bd46c4
Deleted: sha256:e42cc24e8fc51d40dd13284bfaf6f04d52a3859e406929df8ed5889e51ca8159
Deleted: sha256:96e0707818f28da34cb160e28ac98abb05d851e3d12d27b3434601a6805ccdd4
Deleted: sha256:fb3634212649054 1b837f0b49518e884e4537111c933e0e55e0cfeabde02202a
Deleted: sha256:20b6aff657fa996a829a9e9e8099360d137de69d1ec7f4cab5d996f305d835ce
Deleted: sha256:37dc747a118564a8081a97a4430f56255e5f394e36c93dc00ee41b2ec4d7f179
Deleted: sha256:6722f2cffe575c123128b128751c39feadc43948836a0deafbd7bf40c8067203
Deleted: sha256:93d80ac427343fef7b213d17aab0302dc34a5e394edd10b2ae86f3a30d182ca4
Deleted: sha256:7da3459787c37721f80825cc575a5bdde2ee4e6ccbc43b8a89ffe1a587838fc4
Deleted: sha256:8dc11eb4ef5c3c0073e9bb7ed949460d698b544e289a4f441719baa5f3b9c3cf
Deleted: sha256:5b291feac7bdb277c292ef8b16d2795bf74b953a501ca5fa7a11d1d2b9355a50
```

Building a new container

```
./build.sh --name=llm l4t-text-generation\
    tritonserver
```

Currently running docker containers

```
docker ps
```

Pushing your customized docker image

```
docker commit CONTAINER:TAG
docker push CONTAINER:TAG
```

Note:

You mush have an account in `https://hub.docker.com/`

End

https://docs.docker.com/

https://www.jetson-ai-lab.com/