

Crane



**An easy and elegant way to build and manage your
notebook images**

Kaun Lin Huang (@kentwelcome), InfuseAI

Who Am I?
Kaun Lin Huang
a.k.a. Kent Huang
<kent@infuseai.io>
SWE @ InfuseAI
🐙 @kentwelcome



Why



Why we build Crane?

The problem we face

- Building a container image is easy for a developer, but not easy for others
- Writing Dockerfiles from scratch is annoying
 - Should update repo before installing applications
 - Should cleanup cache after installing applications
 - All Update, Install, Cleanup actions should be run in a same command
- Need to manually change the tag when pushing to remote registry

Why we built Crane?

Problem from our Users

- Who are the Users?
 - Data scientist, ML researcher
- Users who aren't familiar with using command-line interface
- Users who don't have strong background knowledge of container technology

Real-World Use Cases

aws-workshop/01_Setup_Depe X

github.com/cyourth-cognonic/aws-workshop/blob/9758252e53b527d546f430b24171d5e4b0a051bf/oreilly_book/01_intro/01_Setup_Dependencies.ipynb

385 lines (385 sloc) | 7.1 KB

<>RawBlame

Setup All Workshop Dependencies

Note: This Notebook Will Take A Few Minutes To Complete.

Please Be Patient.

In []:

```
!python --version
```

In []:

```
!pip list
```

Pip

In []:

```
!pip install --disable-pip-version-check -q pip --upgrade > /dev/null
!pip install --disable-pip-version-check -q wrapt --upgrade > /dev/null
```

AWS CLI and AWS Python SDK (boto3)

In []:

```
!pip install --disable-pip-version-check -q awscli==1.18.216 boto3==1.16.56 botocore==1.19.56
```

SageMaker

In []:

```
!pip install --disable-pip-version-check -q sagemaker==2.29.0
!pip install --disable-pip-version-check -q smdebug==1.0.1
!pip install --disable-pip-version-check -q sagemaker-experiments==0.1.26
```

PyTorch

In []:

```
!conda install -y pytorch==1.6.0 -c pytorch
```

TensorFlow

In []:

```
!pip install --disable-pip-version-check -q tensorflow==2.3.1
```



```

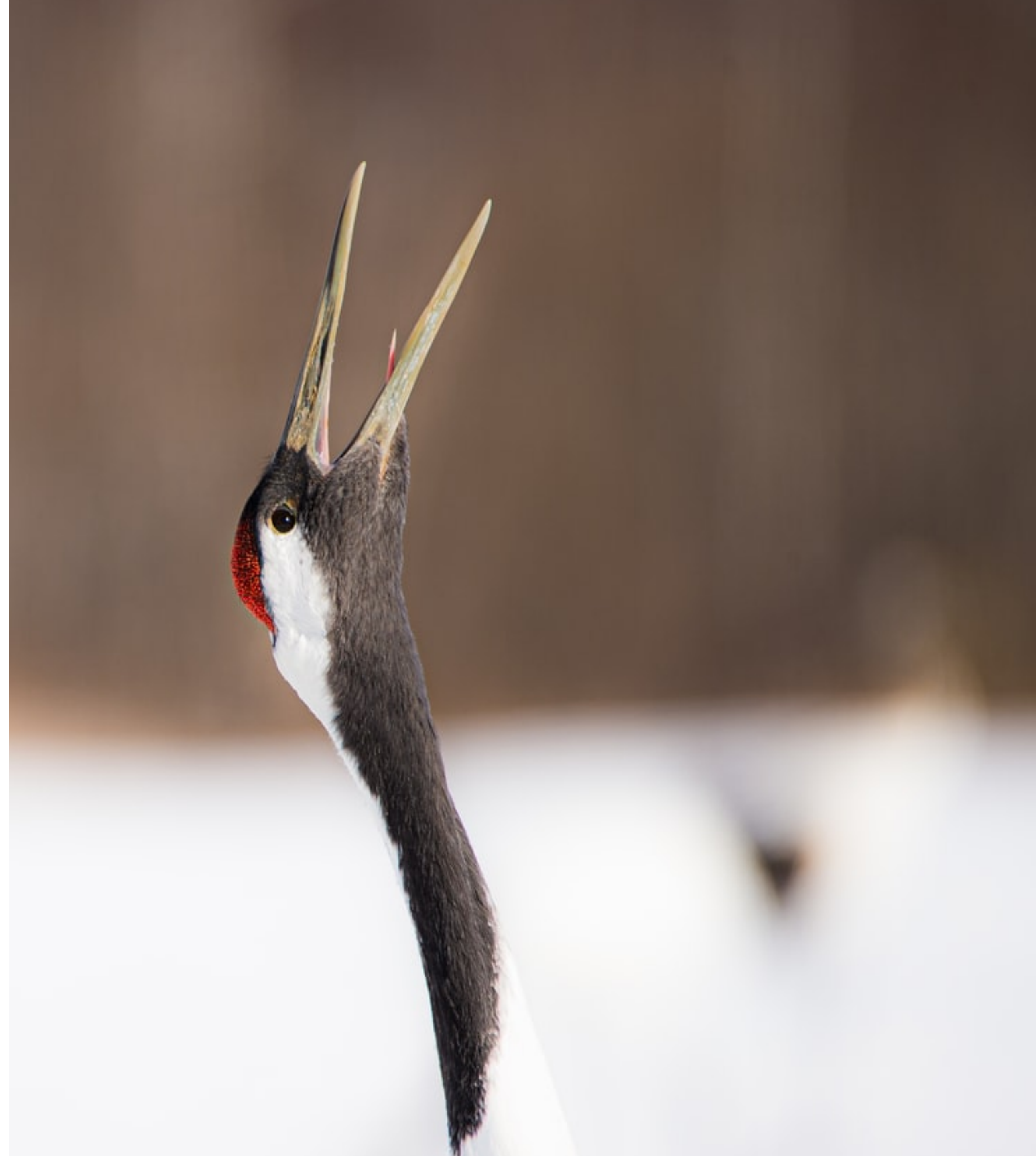
1 FROM nvidia/cuda:10.1-cudnn7-runtime-ubuntu18.04
2 ENV LANG C.UTF-8
3 ENV LC_ALL C.UTF-8
4 ENV KALE_VERSION=0.4.0
5
6 RUN apt-get update && export DEBIAN_FRONTEND=noninteractive && echo "America/Mexico_City" > /etc/timezone && apt-get install -y tzdata
7
8 RUN apt-get update && apt-get install -y \
9     sudo \
10    nano \
11    less \
12    git \
13    wget \
14    curl \
15    python-dev \
16    python3-dev \
17    python3-pip \
18    python3-setuptools && pip3 install --upgrade pip
19 RUN curl -sL https://deb.nodesource.com/setup_13.x | sudo -E bash - && apt-get install nodejs
20 RUN groupadd audiouser
21 RUN useradd audiouser -g audiouser -m -s /bin/bash
22 RUN echo 'audiouser ALL=(ALL) NOPASSWD:ALL' | (EDITOR='tee -a' visudo)
23 RUN echo 'audiouser:audiouser' | chpasswd
24 RUN pip3 install jupyter jupyterlab==1.2.6 --upgrade
25 RUN pip3 install numpy matplotlib pandas seaborn
26 RUN pip3 install dask distributed --upgrade && pip3 install bokeh
27 RUN pip3 install setuptools==46.1.3 tensorflow==2.1.0 librosa==0.7.2 scikit-learn==0.22.2 scikit-image==0.16.2 kubeflow-kale==0.4.0 keras==2.3.1
28 RUN pip3 install torch==1.4.0+cu100 torchvision==0.5.0+cu100 -f https://download.pytorch.org/whl/torch_stable.html
29 USER audiouser
30 RUN jupyter notebook --generate-config && \
31     sed -i "s/#c.NotebookApp.certfile = .*/c.NotebookApp.certfile = u'\/shared_volume\/certs\/fullchain1.pem'/" ~/.jupyter/jupyter_notebook_config.py && \
32     sed -i "s/#c.NotebookApp.keyfile = .*/c.NotebookApp.keyfile = u'\/shared_volume\/certs\/privkey1.pem'/" ~/.jupyter/jupyter_notebook_config.py && \
33     sed -i "s/#c.NotebookApp.password = .*/c.NotebookApp.password = u'sha1:9225c39529b2:3881f9da97d9e3195576e39b70d697992322b4eb'/" ~/.jupyter/jupyter_notebook_config.py && \
34     sed -i 's/#c.NotebookApp.port = .*/c.NotebookApp.port = 9999/' ~/.jupyter/jupyter_notebook_config.py
35 RUN sudo jupyter labextension install kubeflow-kale-launcher
36 VOLUME ["/shared_volume"]
37 ENTRYPOINT ["/usr/local/bin/jupyter", "lab", "--ip=0.0.0.0", "--no-browser"]
38
39

```

Better clean up the apt cache by removing `/var/lib/apt/lists` to reduces the image size

Better to add `--no-cache-dir` when pip install

What



What's Crane

- A minimalist container image builder
- Desktop application for Windows and macOS
- Extend an existing container image with custom apt/conda/pip packages without writing any Dockerfile
- Manage and push local container images to remote registries (ex. Dockerhub, AWS ECR)
- Inspect a container image with Image Layer information

How



How Crane Work

- As a user-interface of the local Docker Engine
 - Generates Dockerfile and builds the container image
 - Lists the container images
 - Inspects the selected container images by the Image Layer
 - Pushes images to remote registries
 - Dockerhub => Docker HUB API
 - AWS ECR => AWS SDK

Crane

+

Crane / Build a new Image

* Image Name

* Base Image

jupyter/base-notebook:latest

Description

Label

jupyter ×

homework ×

+ New Label

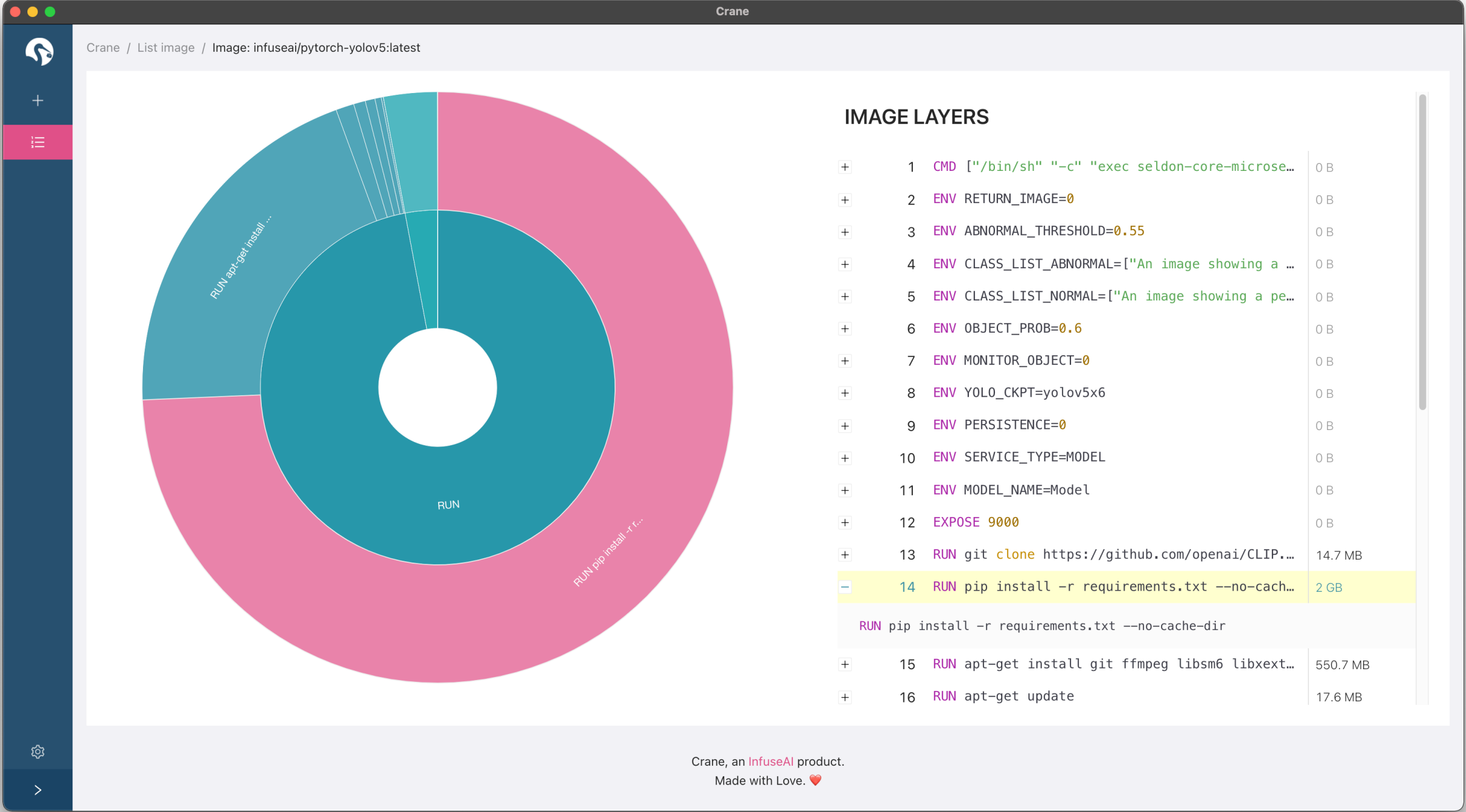
apt

conda

one package per line. e.g.,
package1
package2

one package per line. e.g.,
package1
package2

Crane, an InfuseAI product.
Made with Love.



Crane

+

Crane / List Docker Image

LOCAL

REMOTE

Choose a remote warehouse: DockerHub

REFRESH

	TAG	STATUS	SIZE	LAST UPDATED	
-	infuseaidev/ubuntu	PUBLIC	-	1 day ago	
	infuseaidev/ubuntu:latest	ACTIVE	28.6 MB	1 day ago	
-	infuseaidev/primehub-remote-deploy	PUBLIC	-	1 day ago	Add to PrimeHub
	infuseaidev/primehub-remote-deploy:latest	ACTIVE	472.9 MB	1 day ago	ADD
+	infuseaidev/alpine	PUBLIC	-	1 day ago	
+	infuseaidev/jupyter-notebook-aws	PUBLIC	-	3 weeks ago	
+	infuseaidev/cranetest	PUBLIC	-	1 month ago	
+	infuseaidev/crane	PUBLIC	-	1 month ago	
+	infuseaidev/primehub-notebook	PUBLIC	-	1 month ago	

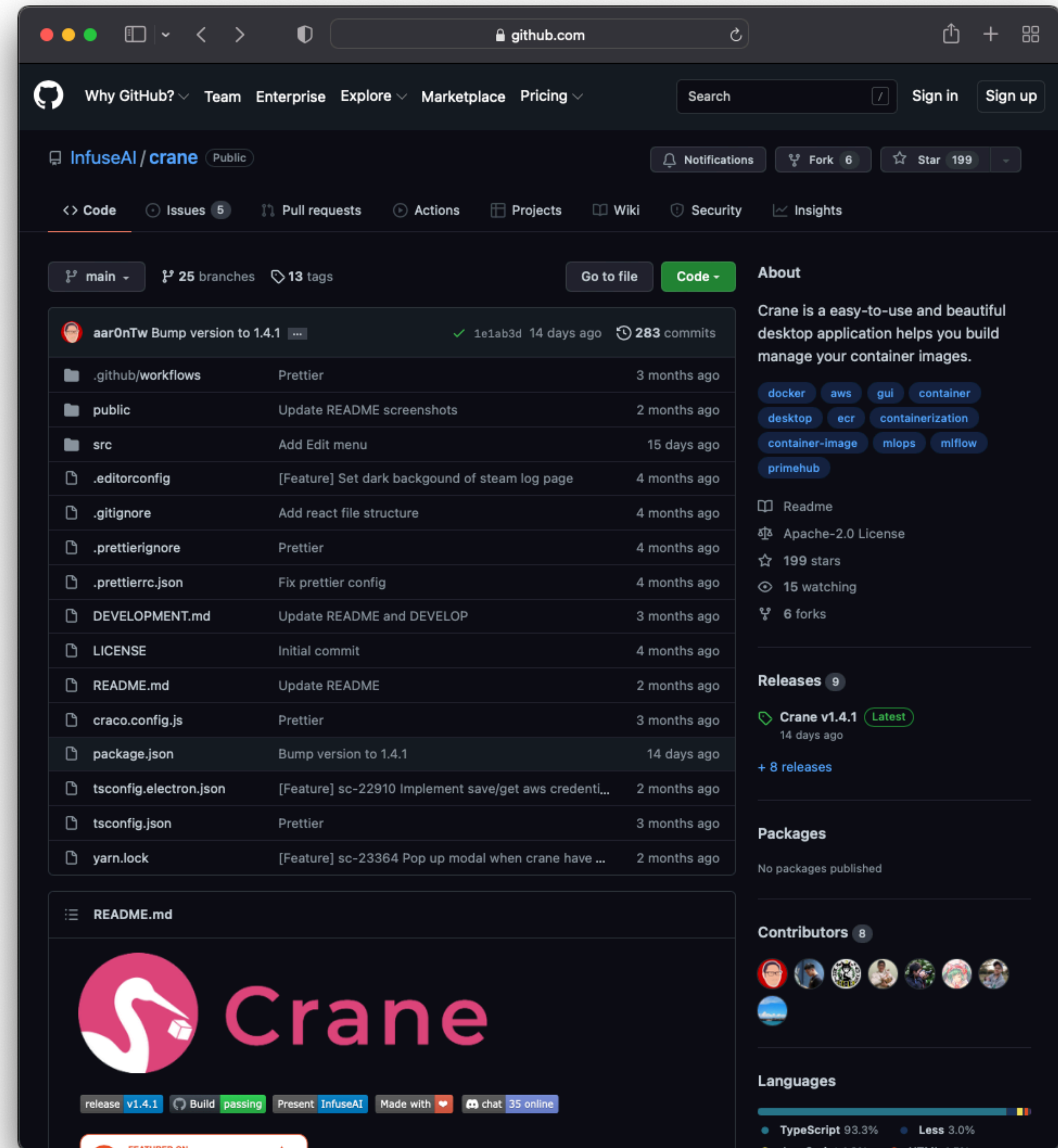
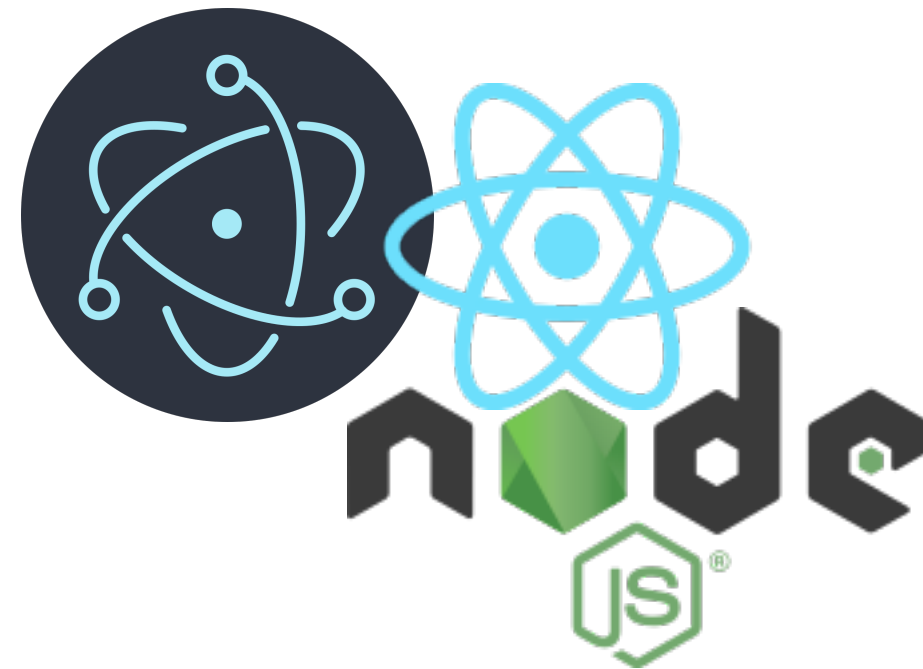
Crane, an InfuseAI product.
Made with Love.

Demo

Inside the Crane

Technical Detail

- Desktop application created by Electron
- FrontEnd: React
- BackEnd: Node.js



How to let Node.js access the Docker Engine?

Node.js + Docker

- Dockerode
 - <https://www.npmjs.com/package/dockerode>
- Via Docker Engine API

```
1 curl --unix-socket /var/run/docker.sock "http://localhost/images/json"
```

```
1 curl --unix-socket /var/run/docker.sock "http://localhost/build" \  
2 -X POST --data-binary "@./dockerfile.tar"
```

```
Darwin Kents-Mac-mini.local 21.1.0 arm64
11:52 up 25 days, 14:05, 2 users, load averages: 7.99 5.35 4.76
dev-eks-kent kent ~ $ curl --unix-socket /var/run/docker.sock "http://localhost/images/json" | jq
Fri Jan 28 11:52:28 2022
dev-eks-kent kent ~ $
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           Dload  Upload   Total     Spent    Left     Speed
100 18900    0 18900    0    0    636k      0 --:--:-- --:--:-- --:--:-- 659k
[
  {
    "Containers": -1,
    "Created": 1643299852,
    "Id": "sha256:fdc8c1370c03d4359586e4aad0439468a12ca5272d8a513ef3c0b0db2608c555",
    "Labels": {
      "com.nvidia.cudnn.version": "7.6.5.32",
      "crane.description": "Demo for FOSDEM 2022",
      "crane.labels": "Cuda-10.1,cudnn7,ubuntu 18.04",
      "maintainer": "NVIDIA CORPORATION <cuda@nvidia.com>"
    },
    "ParentId": "sha256:cb90ee16b8223611a7a8a309e278b4fad8940e5e47bc4dd1ebf976867ed96e45",
    "RepoDigests": null,
    "RepoTags": [
      "nvidia-cuda-notebook:latest"
    ],
    "SharedSize": -1,
    "Size": 1625802521,
    "VirtualSize": 1625802521
  },
  {
    "Containers": -1,
    "Created": 1641875265,
    "Id": "sha256:015dc49dbde1e11ac9adc04789b0795a29538da93ff5e19acb1c7fd9c7c089ed",
    "Labels": {
      "maintainer": "Bitnami <containers@bitnami.com>"
    },
    "ParentId": "",
    "RepoDigests": [
      "bitnami/zookeeper@sha256:f6e57bab4d08e719e4ecd55de02a7279d4acbf5dea211b83da179ccb09f5d0b6"
    ],
    "RepoTags": [
      "bitnami/zookeeper:3.7"
    ],
    "SharedSize": -1,
    "Size": 468162052,
    "VirtualSize": 468162052
  },
  {
    "Containers": -1,
    "Created": 1641801613,
    "Id": "sha256:df0a65f93376cfe4fb356ac315fdfe50e224c5a221da1eb6fe2b49f99b3db9e9",
    "Labels": {
      "maintainer": "Bitnami <containers@bitnami.com>"
    },
    "ParentId": "",
    "RepoDigests": [
      "bitnami/kafka@sha256:37fdf49422fd961093f9026ac5a1c42b9fc75a9103fd962dba37f89934cc1853"
    ],
    "RepoTags": [
      "bitnami/kafka:3"
    ],
    "SharedSize": -1,
    "Size": 614721186,
    "VirtualSize": 614721186
  },
  {
    "Containers": -1,
    "Created": 1641507674,
    "Id": "sha256:554b9758e426241b89c6d4e8cd9b67ced3082eacee0e2c8020c6e305cfaa1bc4",
    "Labels": {}
  }
]
```

Future Roadmap

- More ways to customize the container image
- Ability to download the generated Dockerfile
- Support more remote registries
 - Private registry (ex. Docker registry, Harbor, etc...)
 - GCR Registry, Azure Container Registry

Summary

- Crane is minimalist container image builder
- Our goal is to enable more users to use containers
 - Focus on ML researchers & Data scientists
- Welcome to give feedback in our GitHub repo
 - <https://github.com/InfuseAI/crane>
- Join our Discord
 - <https://discord.gg/ZE8pQ8gRWy>

Reference

- https://docs.docker.com/develop/develop-images/dockerfile_best-practices
- https://github.com/cyourth-cognonic/aws-workshop/blob/9758252e53b527d546f430b24171d5e4b0a051bf/oreilly_book/01_intro/01_Setup_Dependencies.ipynb
- https://github.com/CONABIO/kube_sipecam/blob/102777676f3d72380538e9e8257e2ffdd5cdfbcd/dockerfiles/audio/tf_kale/0.4.0_2.1.0/Dockerfile
- <https://docs.docker.com/engine/api/sdk/examples/>

Thanks

Q&A