機械学習のためのdocker

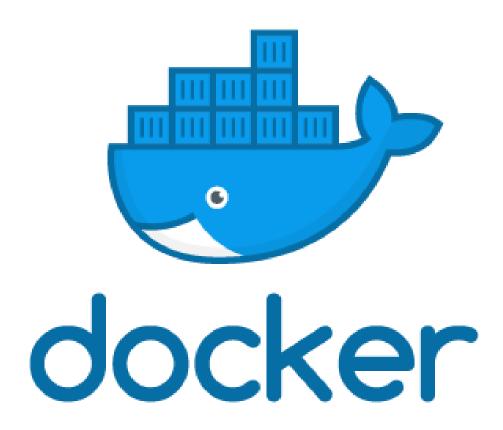
2020/10/14

木南 貴志

1.docker

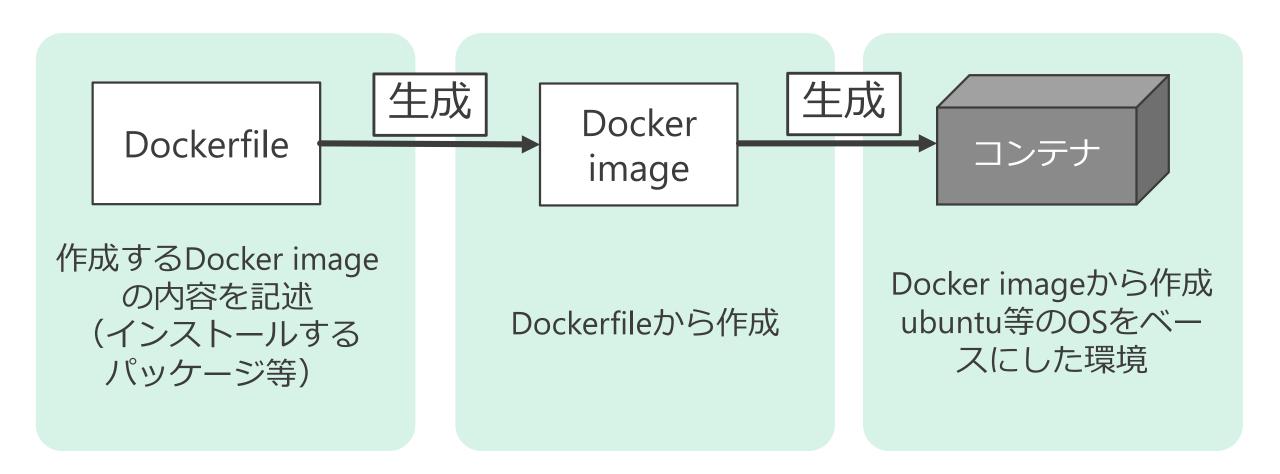
docker

docker: 一つのPC内に仮想環境(のようなもの)を作成するツール



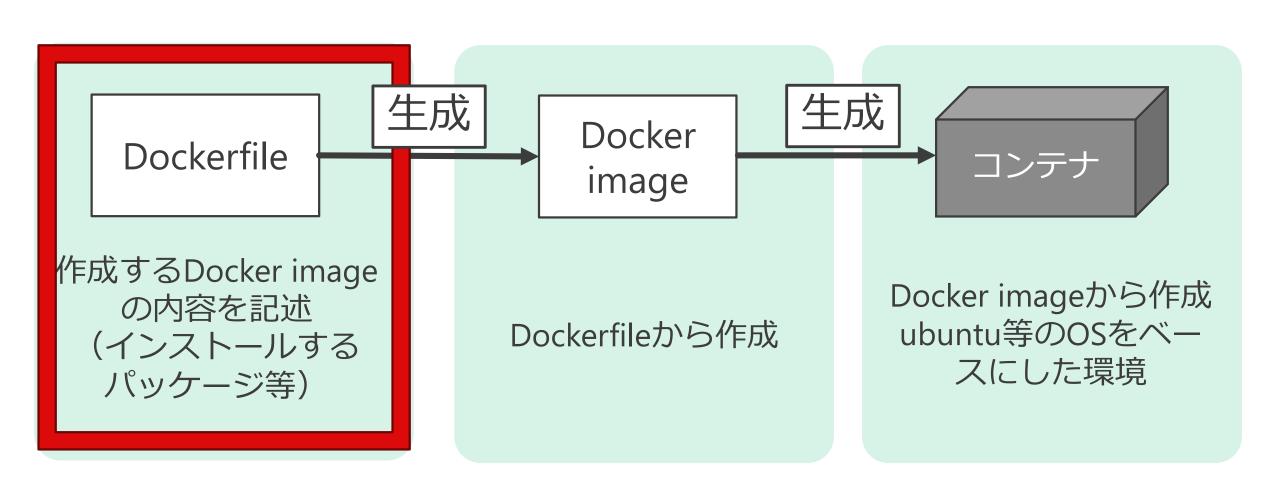
コンテナ作成 流れ

コンテナ作成のおおまかな流れ



コンテナ作成 流れ

コンテナ作成のおおまかな流れ



2.Dockerfile

Dockerfile

Dockerfile

環境構築のための 情報を記述したもの

Dockerfileの名前は 「Dockerfile」にすること ※ buildの際「Docker file」と

いう名前のファイルを探して buildするため

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget
https://repo.anaconda.com/archive/Anaconda3-2020.02-
Linux-x86_64.sh && /
  sh Anaconda3-2020.02-Linux-x86_64.sh -b -p
/opt/anaconda3 && /
  rm -f Anaconda3-2020.02-Linux-x86_64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90==5.3.0 /
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

instruction

instruction: Dockerfileに記述する際のコマンドのようなもの

FROM <base>Docker-image>

説明:ベースのimage(主にubuntu等のOS)を指定

RUN < command >

説明: FROMで指定したOSで使用できるコマンドを実行

ENV <path>

説明: pathを通す(インストールしたパッケージ等の)

CMD <command>

説明:コンテナ生成時に実行されるコマンドを指定

FROM

FROM < Docker image>:ベースとするDocker imageを記述(だいたいはOS)

→ このOSの上にパッケージインストール等して環境を構築

FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04

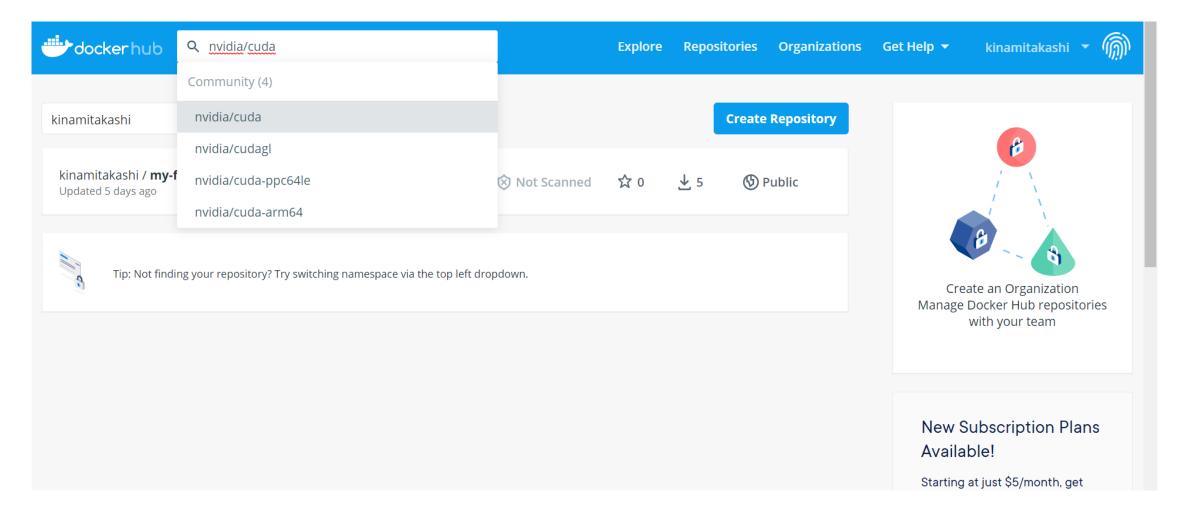


Docker hub(Docker imageを管理するサイト) からベースimageを持ってくる

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget
https://repo.anaconda.com/archive/Anaconda3-
2020.02-Linux-x86 64.sh && /
  sh Anaconda3-2020.02-Linux-x86 64.sh -b -p
/opt/anaconda3 && /
  rm -f Anaconda3-2020.02-Linux-x86 64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90==5.3.0 /
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

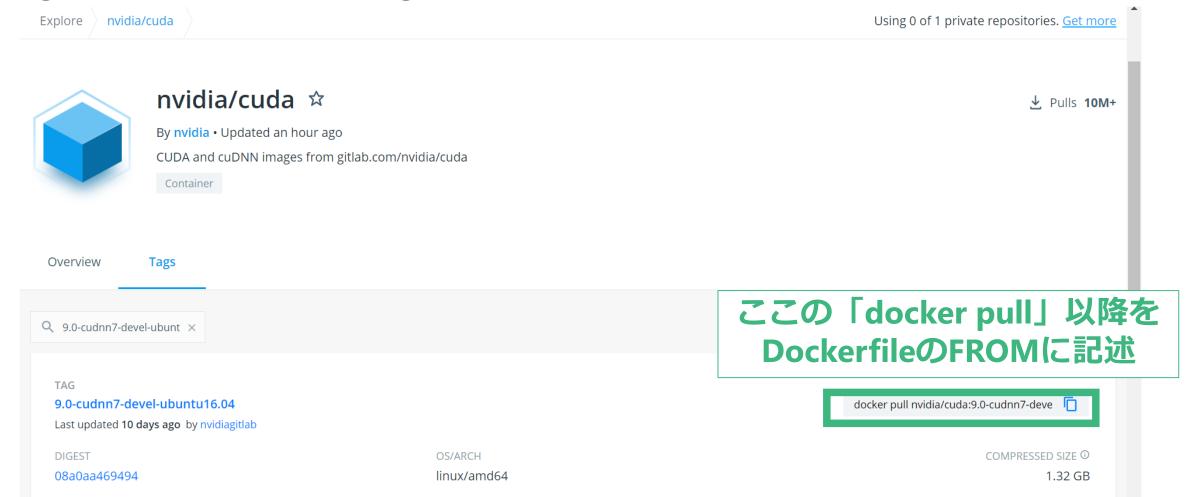
ベースimageを探す

docker hubで使用したいDocker imageを検索



ベースimageを探す

Tagsから使用したいimageのバージョンを確認



RUN

RUN <command>:ベースOSで使用できるコマンドを記述

```
RUN apt-get update && apt-get install -y /
sudo /
wget /
vim
```

ベースOSで使用できるコマンド (linux系ならcdやls等)を記述

必要となるパッケージ等をインストールする等 して環境を設定

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget https://repo.anaconda.com/archive/Anaconda3-
2020.02-Linux-x86 64.sh && /
  sh Anaconda3-2020.02-Linux-x86 64.sh -b -p
/opt/anaconda3 && /
  rm -f Anaconda3-2020.02-Linux-x86 64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90 = = 5.3.0 /
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

WORKDIR

WORKDIR <path>:作業ディレクトリを変更する

WORKDIR /opt

RUN等でコマンドを実行する場所を変更

注意

下記のような場合「hoge」は「/opt」ではなく「/」に作成される

WORKDIR /
RUN cd /opt
RUN mkdir hoge

FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04 RUN apt-get update && apt-get install -y / sudo / wget / vim WORKDIR /opt RUN wget https://repo.anaconda.com/archive/Anaconda3-2020.02-Linux-x86_64.sh && / sh Anaconda3-2020.02-Linux-x86 64.sh -b -p /opt/anaconda3 && / rm -f Anaconda3-2020.02-Linux-x86 64.sh ENV PATH /opt/anaconda3/bin:\$PATH RUN pip install --upgrade pip && pip install / cupy-cuda90 = = 5.3.0 /chainer==5.3.0 WORKDIR / CMD ["/bin/bash"]

RUN cd /opt && mkdir hoge で一度に実行すれば「/opt」に作成される

ENV

ENV <path>:パスを通す

ENV PATH /opt/anaconda3/bin:\$PATH

パッケージ等がすぐに使用できるようにパスを通しておく

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget https://repo.anaconda.com/archive/Anaconda3-
2020.02-Linux-x86 64.sh && /
  sh Anaconda3-2020.02-Linux-x86 64.sh -b -p
/opt/anaconda3 && /
  rm -f Anaconda3-2020.02-Linux-x86 64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90==5.3.0/
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

CMD

CMD ["command"]: コンテナ作成時に実行するコマンドを指定

CMD ["/bin/bash"]

コンテナを作成する際に実行. 原則最後の行に記述.

/bin/bash の場合、シェルを操作できるようになる

※特に理由がない場合、/bin/bashを設定することが多い

このコマンドは下記の様にコンテナ作成時に上書き可能

\$docker run --gpus all cuda9.0cnn7chainer5.3.0 nvidia-smi

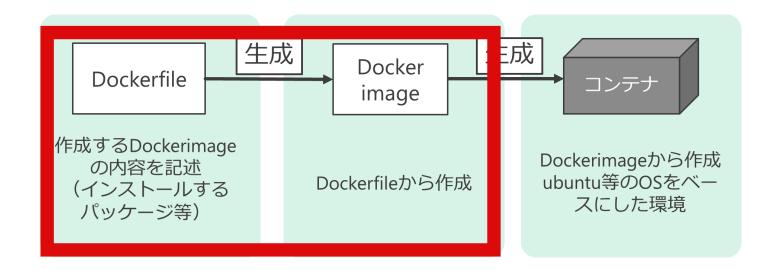
※ /bin/bash の代わりに nvidia-smi が実行される

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  waet /
  vim
WORKDIR /opt
RUN wget https://repo.anaconda.com/archive/Anaconda3-
2020.02-Linux-x86 64.sh && /
  sh Anaconda3-2020.02-Linux-x86 64.sh -b -p
/opt/anaconda3 && /
  rm -f Anaconda3-2020.02-Linux-x86 64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90 = = 5.3.0 /
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

docker build

Dockerfile → Docker imageを生成

\$docker build -t <Docker imageの名称(任意)> <Dockerfileのある場所(相対パス)>



例)Docker fileがある場所が現在地(.),生成するDocker imageの名前をcuda9.0chainer5.3.0とする場合

\$docker build -t cuda9.0cudnn7chainer5.3.0.

Docker image

Docker imageはDocker layerを重ねたもの(1layer=1instruction)

Docker image

Docker-layer4

Docker-layer3

Docker-layer2

Docker-layer1

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget https://repo.anaconda.com/archive/Anaconda3-
2020.02-Linux-x86 64.sh && /
  sh Anaconda3-2020.02-Linux-x86_64.sh -b -p /opt/anaconda3
88/
  rm -f Anaconda3-2020.02-Linux-x86_64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90==5.3.0 /
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

Docker image

Docker imageはDocker layerを重ねたもの(1layer=1instruction)

```
Docker image
Docker-layer4
Docker-layer3
Docker-layer2
Docker-layer1
```

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget https://repo.anaconda.com/archive/Anaconda3-
2020.02-Linux-x86_64.sh && /
  sh Anaconda3-2020.02-Linux-x86_64.sh -b -p /opt/anaconda3
88/
  rm -f Anaconda3-2020.02-Linux-x86_64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90==5.3.0/
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

Docker layerを増やすメリット

layer毎にbuildされるため, build時間が短縮

Docker-layer2

Docker-layer1

1回目

```
FROM ubuntu:16.04
RUN apt-get update && apt-get install -y /
sudo /
wget /
vim
```

Docker-layer3

Docker-layer2

Docker-layer1

2回目

```
FROM ubuntu:16.04
RUN apt-get update && apt-get install -y /
sudo /
wget /
vim
RUN apt-get install -y /
python
```

layer1, 2はbuild済のためlayer3のみbuild

Docker layerを増やすデメリット

instructionが増えるとdocker imageの容量も増加

```
RUN apt-get update && apt-get install -y / sudo / wget / vim
```

build後のimage file

→ 容量小

→ 谷重小

```
RUN apt-get update
RUN apt-get install –y sudo
RUN apt-get install –y wget
RUN apt-get install –y vim
```

build後のimage file

→ 容量大

同じ処理内容だがbuild後のimage fileの容量は下例の方が大

環境構築の進め方

環境構築中は, instructionを1行ずつ追加(buildの時間削減)

➡ 構築し終わったらDocker imageの容量を小さくするため, instructionをまとめる

RUN apt-get update

RUN apt-get install -y sudo

RUN apt-get install -y wget

RUN apt-get install -y vim

RUN apt-get update && apt-get install -y /

sudo /

wget /

vim

環境構築中

➡instructionを多用

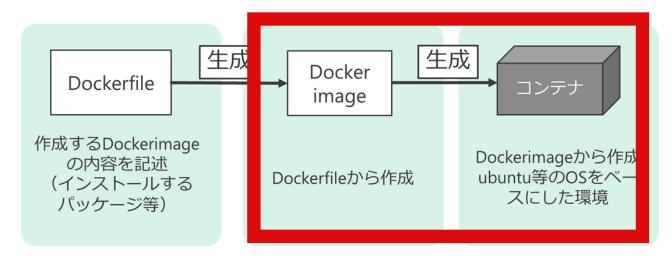
環境構築終了後

→ instructionを最小限にまとめる

docker run

Docker image → containerを生成(+コンテナ内に入る)

\$docker run -it --name=<コンテナの名称(任意)> <Docker image> <コンテナ起動時に実行したいコマンド>

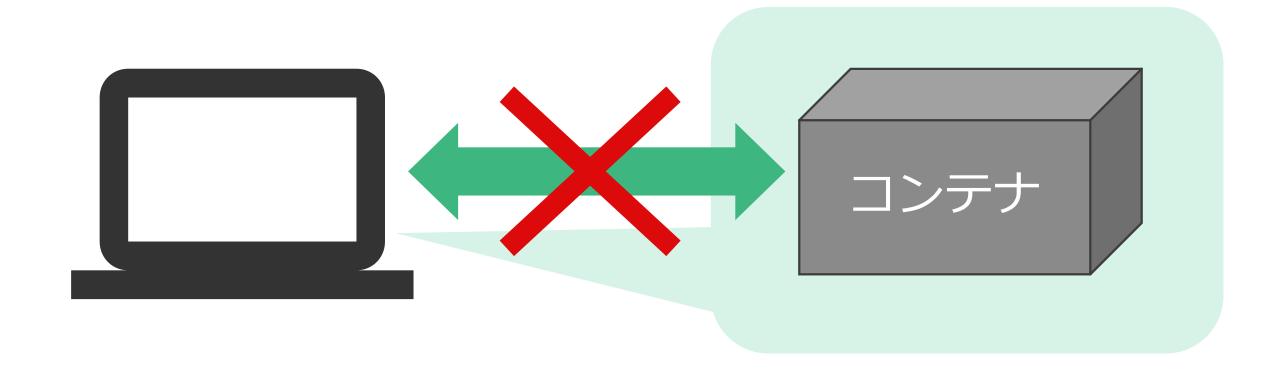


例)Docker image「cuda9.0cnn7chainer5.3.0」からコンテナ「cuda90cudnn7chainer530」を生成

\$docker run -it --name=cuda90cudnn7chainer530 cuda9.0cudnn7chainer5.3.0 bash

ホスト↔コンテナでファイル共有

通常,ホスト(自分のPC) →コンテナの間でファイルの共有は不可

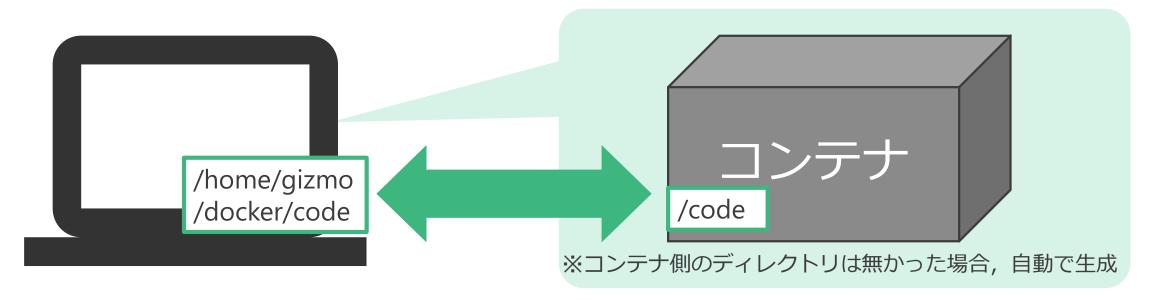


ホスト→コンテナでファイル共有

-vオプションでディレクトリのマウントが可能

\$docker run -it -v /home/gizmo/docker/code:/code cuda9.0cudnn7chainer5.3.0

-v <ホスト側のマウントしたいディレクトリまでのパス>:<コンテナ側のパス>

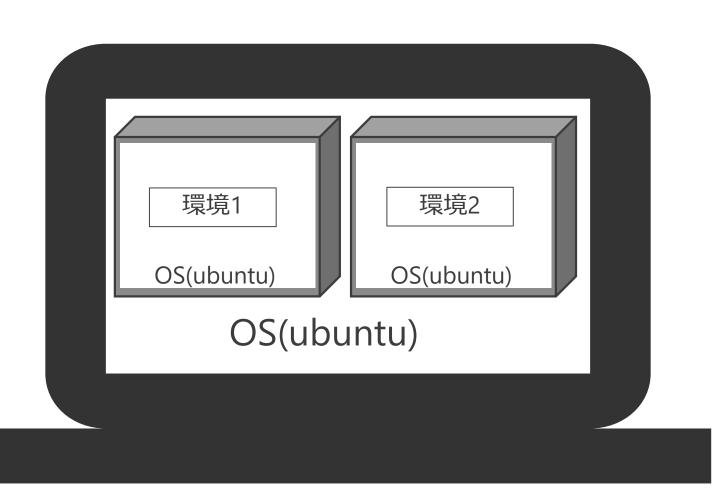


ホストの「/home/gizmo/docker/code」内のファイルがコンテナの「/code」に反映 ※参照しているだけなのでホスト側でファイルの追加・編集するとコンテナ側にも反映(逆も)

2.機械学習を行うための docker

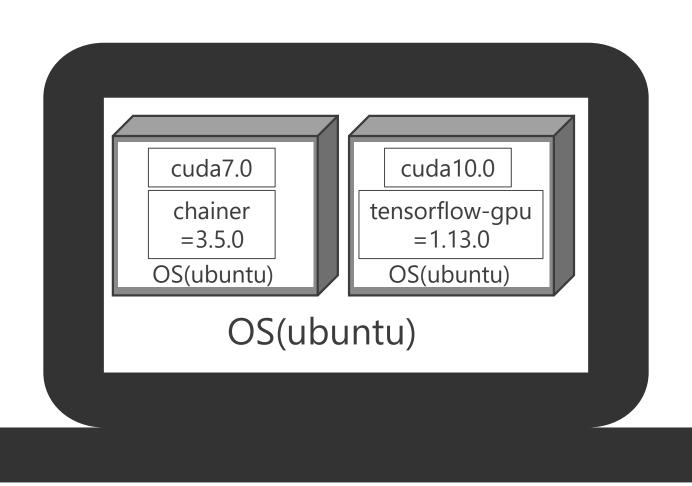
機械学習でコンテナを使う利点①

各コンテナで異なる環境を構築可能



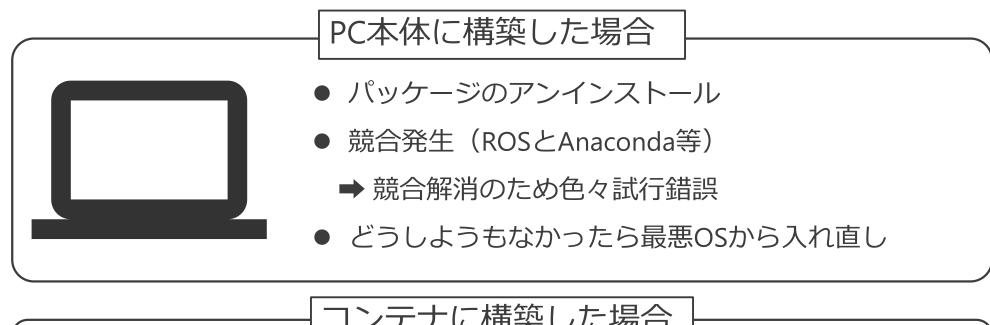
機械学習でコンテナを使う利点①

例) コンテナ1にcuda7.0, コンテナ2にcuda10.0の環境を構築



機械学習でコンテナを使う利点②

間違った環境構築をした場合 ➡ コンテナから抜けるだけでOK



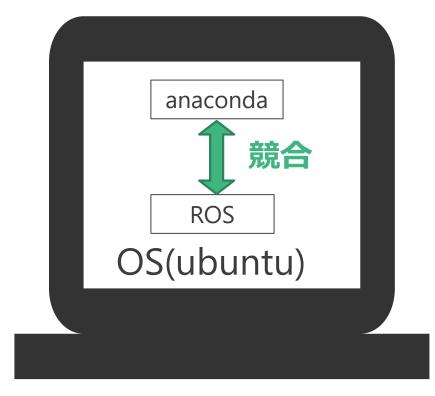
コンテナに構築した場合

コンテナから抜ける. 以上

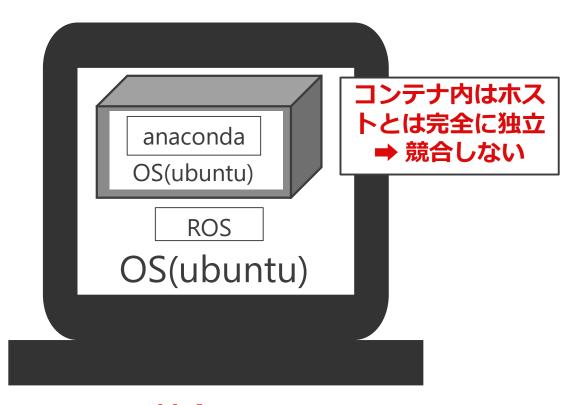
機械学習でコンテナを使う利点③

コンテナ内は完全に独立した環境 → 競合の回避

(ただし,設定すればホスト↔コンテナ間,コンテナ↔コンテナ間で共有することも可)



競合の可能性あり



競合しない

ホストとコンテナの関係性

機械学習の場合のイメージ

docker, nvidia-container-toolkitのみ ホスト側にインストール

docker

nvidia-containar-toolkit

NVIDIAドライバー

それ以外(cuda, 機械学習フレーム ワーク, python等)全てをコンテナ 内のOS上にインストール

cuda 機械学習フレームワーク その他必要なパッケージ OS(ubuntu)

手順

- 1. Dockerをインストール
- 2. 自分のPCに搭載されているGPUの, NVIDIA-driver(基本最新でOK) を インストール
- 3. nvidia-container-toolkitをインストール(1.~3.は初回のみ)
- 4. Dockerfileを記述
- 5. Dockerfile → image-fileを作成(ビルド)
- 6. image-file → containerを作成(container =実際に作業する環境)
- ※環境構築が完了するまで、4.~6.を繰り返す
- 7. 環境構築が完了したら, container内で作業(プログラム実行等)

Dockerをインストール

UbuntuにDockerEngineをインストールする

https://docs.docker.com/engine/install/ubuntu/

※Dockerは頻繁に更新されるため、上記の公式のページを参考にして

インストールした方が無難

dockerをsudoなしでログイン

https://qiita.com/DQNEO/items/da5df074c48b012152ee

NVIDIA driver インストール

\$ Ispci | grep -i nvidia

自分のPCに搭載のGPUを調べる

- ・調べたGPUに対応するNVIDIA driverの最新verを確認 https://www.nvidia.co.jp/Download/index.aspx?lang=jp
- ・調べたGPUに対応するNVIDIA driverの最新verのインストールファイルをダウンロード https://www.nvidia.co.jp/Download/index.aspx?lang=jp

\$ chmod +x <ダウンロードしたファイル>

実行できるように

\$ sudo ./<ダウンロードしたファイル>

ドライバインストール

\$ reboot

再起動

\$ nvidia-smi

ドライバがインストールされているか確認

NVIDIA driver インストール

※ nvidia-dimが使用中というエラーが出たら以降を実行

\$ systemctl isolate <u>multi-user.target</u>

CUIモードに変更

CUIモードになったら alt+F1 → ユーザ名、パスワードを入力でログイン

\$ modprobe -r nvidia-drm

無効化

\$ modprobe -r nvidia-modeset

無効化

\$ sudo ./<ダウンロードしたファイル>

ドライバインストール

\$ reboot

再起動

\$ nvidia-smi

ドライバがインストールされているか確認

nvidia-driver 対応CUDA確認

```
$ nvidia-smi
```

GPUの最新driverがCUDAのどのverまで対応しているか確認

```
CUDA 10.0以下まで対応
Tue Oct 22 15:46:59 2019
                      Driver Version: 410.104
                                                CUDA Version: 10.0
 NVIDIA-SMI 410.104
                Persistence-M| Bus-Id
                                          Disp.A | Volatile Uncorr. ECC
     Temp Perf Pwr:Usage/Cap| Memory-Usage | GPU-Util Compute M.
   0 GeForce RTX 2080
                       Off | 00000000:01:00.0 On |
                  24W / 225W | 5409MiB / 7949MiB |
                                                              Default
                                                           GPU Memory
 Processes:
  GPU
                                                           Usage
                Type
          1440
                       /usr/lib/xorg/Xorg
                                                               297MiB
                   G /usr/bin/gnome-shell
          1652
                                                               199MiB
                       /usr/lib/firefox/firefox
          2920
```

最新ドライバーが「使用したいCUDA」に対応していない場合GPUを新しいものに買い換える必要あり

nvidia-container-toolkitをインストール

nvidia-containar-toolkitをインストール

https://github.com/NVIDIA/nvidia-docker

※Dockerは頻繁に更新されるため、上記の公式のページを参考にして

インストールした方が無難

機械学習の際のDockerfileの主な構成

```
FROM nvidia/cuda:9.0-cudnn7-devel-ubuntu16.04
RUN apt-get update && apt-get install -y /
  sudo /
  wget /
  vim
WORKDIR /opt
RUN wget
https://repo.anaconda.com/archive/Anaconda3-2020.02-
Linux-x86 64.sh && /
  sh Anaconda3-2020.02-Linux-x86 64.sh -b -p
/opt/anaconda3 && /
  rm -f Anaconda3-2020.02-Linux-x86_64.sh
ENV PATH /opt/anaconda3/bin:$PATH
RUN pip install --upgrade pip && pip install /
cupy-cuda90==5.3.0 /
  chainer==5.3.0
WORKDIR /
CMD ["/bin/bash"]
```

使用したいubuntu, cuda, cudnnのverを指定

anacondaのインストール

https://repo.anaconda.com/archive で検索して任意のverのanacondaをダウンロード&インストールが終わったらshファイルは削除これで、python関連(numpy, matplotlib等)はほとんどインストールされる

使用したい機械学習のフレームワークと verを指定してインストール

Dockerfile作成の際の注意点

docker buildの際は、キーボードによる入力不可

→ キーボード入力をしなくていいような工夫が必要

```
RUN apt-get update && apt-get install -y /
sudo /
wget /
vim
```

-yをオプションに指定 ➡ キーボード入力を要求されたら全て「y」で返答

```
RUN wget https://repo.anaconda.com/archive/Anaconda3-2020.02-Linux-x86_64.sh && / sh Anaconda3-2020.02-Linux-x86_64.sh -b -p /opt/anaconda3 && / rm -f Anaconda3-2020.02-Linux-x86_64.sh
```

-bをオプションに指定 **→** バッチモード(キーボード入力不要モード)でインストールを実行 ※ -bオプションはshファイル共有のものではないため注意

機械学習におけるコンテナ作成の流れ

基本的に、コンテナ作成 → 破壊を繰り返すことが多い

▶ Dockerfile編集

Dockerfile → Docker image生成

\$docker build -t <Docker imageの名称(任意)> <Docker fileのある場所(相対パス)>

Dockerimage → コンテナ生成(runの際に--rmをオプションにつけるとexit時にコンテナが自動で削除) ◆

\$docker run --gpus all -it --rm --name=<コンテナの名称(任意)> -v <ホスト側のマウントしたいディレクトリまでの絶対パス>:<コンテナ側のディレクトリまでの絶対パス> <Dockerimage> <コンテナ起動時に実行したいコマンド>

環境構築途中

環境構築済

上手く環境構築できなかった または,環境構築に必要な パッケージ等判明

→ Dockerfileに内容を反映



作業開始(プログラム実行など)

コンテナから抜ける 「wit

exit

作業終了 (マウントしておけば, 追加・編集したファイル 等はホスト側に残る)

コンテナ作成の流れ 例)

基本的に, コンテナ作成 → 破壊を繰り返すことが多い

▶ Dockerfile編集

Dockerfile → Docker-image生成

\$docker build -t cuda9.0cudnn7chainer5.3.0.

Dockerimage → コンテナ生成(runの際に--rmをオプションにつけるとexit時にコンテナが自動で削除) ◆

\$docker run --gpus all -it --rm -name==cuda90cudnn7chainer530 -v /home/gizmo/docker/files:/filescuda9.0cudnn7chainer5.3.0 bash

環境構築途中

環境構築済

上手く環境構築できなかった または,環境構築に必要な パッケージ等判明

→ Dockerfileに内容を反映



作業開始(プログラム実行など)

コンテナから抜ける

exit

作業終了 (マウントしておけば, 追加・編集したファイル 等はホスト側に残る)

補足 機械学習のdocker run オプション

docker runで使用するオプションを説明

\$docker run --gpus all -it --rm -name==cuda90cudnn7 -v /home/gizmo/docker/files:/filescuda9.0cudnn7chainer5.3.0 bash

- -it: きれいに表示する. このコマンドがないと上手く表示されない(詳細は省略)
- -v <host>:<container> : ホストのディレクトリとコンテナのディレクトリをマウントする.
- ※ コンテナ内に, 指定したディレクトリが存在しない場合は自動でディレクトリが生成される
- --gpus:ホスト側のGPUをコンテナ内で使用. allを指定すると(ホストに複数のGPUが搭載されている場合)全てのGPUをコンテナで使用
- --rm:exitでコンテナから抜けた後に、コンテナが自動で削除される

コンテナ内でのGPUの認識

\$docker run --gpus all (以下略)

でコンテナを起動したときにエラーが発生した場合, NVIDIAドライバ, nvidia-containar-toolkitが正しくインストールされていない可能性あり

ホストとコンテナ内の両方で

\$nvidia-smi

を実行した時,ホストとコンテナ内で内容が異なっていた場合(CUDAのバージョンにERRと表示されている等)も同様にドライバのインストールができていない可能性あり

3.dockerのコマンド

dockerで使用することの多いコマンド

(Dockerimage, containarを指定する場合, IDか名前のどちらかを指定)

\$ docker build <path_to_Dockerfile>

Dockerfile → Dockerimageを生成

options: -t: Dockerimageに名前をつける

-f < Dockerfilename>: Dockerfileに「Dockerfile」以外の名前をつけている場合

\$ docker run < Dockerimage > < command >

Dockerimageからコンテナを生成

※ <command>を指定しない場合DockerfileのCMDで指定したコマンドを実行

exitしたらコンテナが停止する

options: -v <host>:<path>: 指定したディレクトリをマウント

--name <container_name>: コンテナに名前をつける

-it: きれいに表示する(詳細は省略)

--gpus <number>:ホストのGPUをコンテナ内で使用(allでホストにある全てのGPUをコンテナで使用)

--rm:exitでコンテナから抜けた後にコンテナを削除(指定しない場合,停止はするが削除はされない)

-p <host_port>:<container_port>:ホストのポートとコンテナのポートを接続

-u <user_id><group_id>:ユーザIDとグループIDを指定(通常はrootでログインする)

※ -u \$(id -u) \$(id -g) と指定すれば直接IDを入力しなくて良い

dockerで使用することの多いコマンド

(Dockerimage, containarを指定する場合, IDか名前のどちらかを指定)

\$ docker exec <container> <command>

コンテナに再度入る

※ <command>を指定しない場合DockerfileのCMDで指定したコマンドを実行 options: -it: きれいに表示する (詳細は省略) eixtした場合もコンテナが停止しない.

\$ docker attach < container > < command >

コンテナに再度入る

※ <command>を指定しない場合DockerfileのCMDで指定したコマンドを実行options: -it: きれいに表示する(詳細は省略)

eixtしたらコンテナが停止する

\$ docker restart < container>

停止したコンテナを再アクティブ化

※ 停止している場合exec, attachができないため、停止したコンテナに入りたい場合はこれを実行してから

dockerで使用することの多いコマンド

(Dockerimage, containarを指定する場合, IDか名前のどちらかを指定)

\$ docker commit < container> < new_Dockerimage> コンテナ → Dockerimageを生成

\$ docker tag <befor_Dockerimage> <after_Dockerimage> **Dockerimageの名前変更** Dockerimageの名前は「リポジトリ名:タグ名」

\$ docker push < Dockerimage >

Dockerimageをdocker hubにpush

pushする際は,Dockerimageの名前は「pushする先のリポジトリ名:タグ名」にすること(docker hubにpushする際にDockerimageを参照して,そのリポジトリにpushするため)

dockerで使用することの多いコマンド

(Dockerimage, containarを指定する場合, IDか名前のどちらかを指定)

\$ docker rm < container>

コンテナを削除

停止していないコンテナは削除不可

\$ docker stop < contaner>

コンテナを停止

\$ docker system prune 停止しているコンテナ,名前のついていないdockerimageを全て削除

\$ docker rmi < Dockerimage >

Dockerimageを削除

コンテナ内で使用できるコマンド

\$ exit コンテナから抜ける(抜けた後コンテナ停止)

\$ detatch コンテナから抜ける(抜けた後もコンテナ動作継続)

参考資料

参考サイト(超おすすめ)

https://datawokagaku.com/docker_lecture/

↑セールは行わないとのことなので,期間限定の割引クーポンを使用 して講座を購入することをおすすめします