

Docker for Machine Learning

- First steps – just use image
- More advanced – modify a public image
- More realistic – use (changed) images together

Just use image

Sources for Jupyter Notebooks

- <https://hub.docker.com/u/jupyter>
- <https://github.com/jupyter/docker-stacks>
- <https://github.com/Kaggle/docker-python>
- ...

Let's pick a universal one:

<https://hub.docker.com/r/jupyter/datascience-notebook>

Just use image

Steps

- Pull Image
- Run it

Can be one command

1) Download + Start

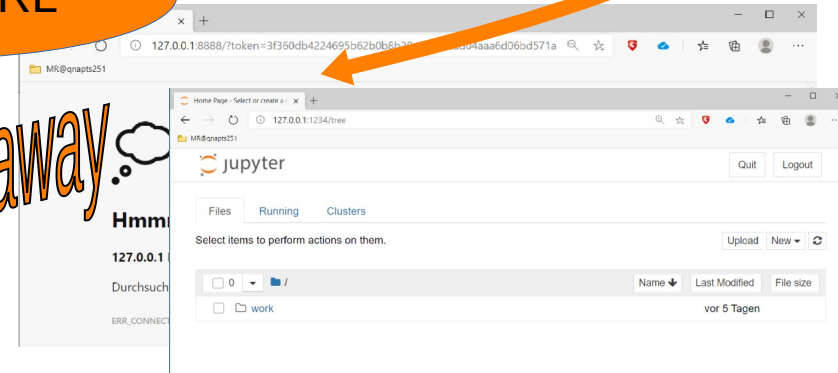
```
> docker run jupyter/datascience-notebook
```

3) Paste URL

2) Copy URL

May or may not work right away

```
PS E:\QSync\Mobile\Infos\Analytics\Big Data\Data Science\Machine Learning\ML DevOps> docker run jupyter/datascience-notebook
Unable to find image 'jupyter/datascience-notebook:latest' locally
latest: Pulling from jupyter/datascience-notebook
4a7291352a9b: Pull complete
140428a6d4bcd: Pull complete
2c2d948710f2: Pull complete
ca78f2d1592f: Pull complete
2a7a2fbed339: Pull complete
d8b4c5bd8e6d: Pull complete
773273f9a979: Pull complete
73968a1cce87: Pull complete
b96c7c3f9e37: Pull complete
3c47e33b3e4d: Pull complete
0f68d404db95: Pull complete
62111c688dc9: Pull complete
1165e39c1e18: Pull complete
0e36996f6dd3: Pull complete
f9c15286a9d3: Pull complete
f79fd95f721: Pull complete
ec79857d12e8: Pull complete
556c3a8978ae: Pull complete
1696bccf21c7: Pull complete
729adc4b79ab: Pull complete
2c1bda956783: Pull complete
821bb95dab97: Pull complete
348dabe791df: Pull complete
2f8ef3706b64: Pull complete
eabb550054b1: Pull complete
38cb299b6e9d: Pull complete
d8528bd28762: Pull complete
Digest: sha256:a1cd0b731489799fc8cc7b451a8a3d6ee6675691f9658a4b0ff856f9c8f9271b
Status: Downloaded newer image for jupyter/datascience-notebook:latest
Executing the command: jupyter notebook
[I 07:45:02.503 NotebookApp] Writing notebook server cookie secret to /home/jovyan/.local/share/jupyter/runtime/notebook_cookie_secret
[I 07:45:02.923 NotebookApp] JupyterLab extension loaded from /opt/conda/lib/python3.8/site-packages/jupyterlab
[I 07:45:02.925 NotebookApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 07:45:02.925 NotebookApp] Serving notebooks from local directory: /home/jovyan
[I 07:45:02.925 NotebookApp] Jupyter Notebook 6.1.6 is running at:
[I 07:45:02.925 NotebookApp] http://cc6c7ca6581b:8888/?token=3f360db4224695b62b0b8b20aea596cad04aaa6d06bd571a
```



Just use image

Troubleshooting

1. Stop running container

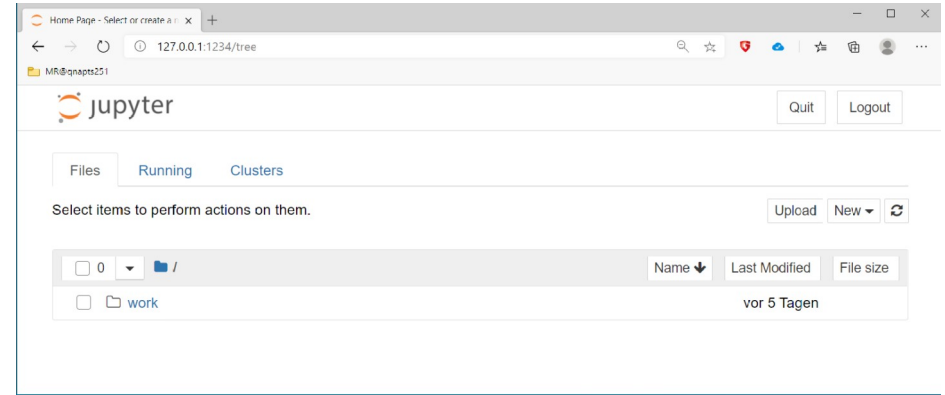
- `<Ctrl+C>`

1. Remove running container

- `docker container rm -f <mycontainer>`

2. Restart container with parameter(s)

- `docker run -p 1234:8888 jupyter/datascience-notebook`



Just use image

Working with Container

- Name container
 - `docker run -p 1234:8888 --name jupyter_ds --rm jupyter/datascience-notebook`
- Use data files (*.csv, etc.)
 - `docker cp some.csv jupyter_ds:/home/jovyan/`
- Persist *.ipynb, etc. outside container
 - `docker run -p 1234:8888 -v C:\Users\Mike_64\Documents\MLOps_data:/home/jovyan --rm --name jupyter_ds jupyter/datascience-notebook`

Modify a Public Image

Create individual Docker image with Dockerfile:

- Pull an image
- Customize it
- Save it to own Docker registry
- Run containers of the modified image

Modify a Public Image

Steps

- 1) Code Dockerfile
- 2) Build image
- 3) Run container off image

```
1 FROM jupyter/datascience-notebook
2
3
4 #base image is Ubuntu (https://github.com/docker-library/official-images/949)
5 #to install, user needs to be root
6 USER root
7
8 #update base packages
9 RUN apt-get update && \
10     apt-get --yes install \
11     apt-utils \
12     vim
13
14 #adding Python packages
15 RUN pip install mysql.connector sqlalchemy
16
17 #preparing Notebook environment
18 RUN mkdir /home/jovyan/.jupyter/
19 COPY jupyter_notebook_config.py /home/jovyan/.jupyter/
20
21 #back from root
22 USER $NB_UID
```

1

```
docker build -t mike_jn .
```

2

```
docker run -p 1234:1234 -v C:\Users\Mike_64\Documents:/home/jovyan --rm --name jupyter_ds mike_jn
```

```
[I 07:50:39.946 NotebookApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 07:50:39.948 NotebookApp] Serving notebooks from local directory: /home/jovyan
[I 07:50:39.948 NotebookApp] Jupyter Notebook 6.1.6 is running at:
[I 07:50:39.948 NotebookApp] http://e60e0fab498:1234/?token=b1e99e5af04b7d7d5f21c1ca8e94bd315abe47a8f3eee8e2
[I 07:50:39.948 NotebookApp] or http://127.0.0.1:1234/?token=b1e99e5af04b7d7d5f21c1ca8e94bd315abe47a8f3eee8e2
[I 07:50:39.948 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 07:50:39.989 NotebookApp]
```

To access the notebook, open this file in a browser:
file:///home/jovyan/.local/share/jupyter/runtime/nbserver-7-open.html
Or copy and paste one of these URLs:
http://e60e0fab498:1234/?token=b1e99e5af04b7d7d5f21c1ca8e94bd315abe47a8f3eee8e2

3

<http://127.0.0.1:1234/?token=b1e99e5af04b7d7d5f21c1ca8e94bd315abe47a8f3eee8e2>

Modify a Public Image

Worth noting:

- Dockerfile name must be „Dockerfile“
- Docker build must be issued in directory of Dockerfile

Modify a Public Image

For best practices:

https://docs.docker.com/develop/develop-images/dockerfile_best-practices/

A good tutorial:

<https://takacsmark.com/dockerfile-tutorial-by-example-dockerfile-best-practices-2018/>

Use (Changed) Images Together

Pull / change multiple image to form „apps“:

- Pull images
- Customize them
- Save them to own Docker registry
- Run them as services with docker-compose

Use (Changed) Images Together

Steps

- 1) Code docker-compose.yml file
- 2) Run docker-compose

```
1 version: '3'
2 services:
3   mysql:
4     image: mike_ms
5     hostname: mike_ms
6     container_name: mike_ms
7     ports:
8       # <Port exposed> : <MySQL Port running inside container>
9       - '2345:3306'
10
11   expose:
12     # Opens port 3306 on the container
13     - '3306'
14     # Where our data will be persisted
15     #restart: always
16
17   environment:
18     MYSQL_DATABASE: 'dsdb'
19     # So you don't have to use root, but you can if you like
20     MYSQL_USER: 'dsdb_user'
21     # You can use whatever password you like
22     MYSQL_PASSWORD: 'dsdb_user_pwd'
23     # Password for root access
24     MYSQL_ROOT_PASSWORD: 'root_pwd'
```

1

docker-compose up

2

```
PS E:\VSync\Media\InfoAnalytics\Big Data\Big Data Science\Machine Learning\UDops>
Creating network 'mlops_data' with the default driver
Creating mike_ms ... done
Attaching to mike_ms, mike_jn
mike_jn    Executing the command: jupyter notebook
mike_ms    2021-01-11T21:47:08.456217Z # [Warning] TIMESTAMP with implicit DEFAULT value is deprecated. Please use --explicit_defaults_for_timestamp server option (see documentation for more details).
mike_ms    2021-01-11T21:47:08.461862Z # [Note] mysqld (mysqld 5.7.16) starting as process 1
mike_ms    2021-01-11T21:47:08.465750Z # [Warning] Setting lower_case_table_names=1 because file system for /var/lib/mysql/ is case insensitive
mike_ms    2021-01-11T21:47:08.467825Z # [Note] InnoDB: PUNCH HOLE support available
mike_ms    2021-01-11T21:47:08.467902Z # [Note] InnoDB: Mutexes and rw_locks use GCC atomic builtins
mike_ms    2021-01-11T21:47:08.467942Z # [Note] InnoDB: Uses event mutexes
mike_ms    2021-01-11T21:47:08.467960Z # [Note] InnoDB: GCC builtin __atomic_thread_fence() is used for memory barrier
mike_ms    2021-01-11T21:47:08.467968Z # [Note] InnoDB: Compressed tables use zlib 1.2.3
mike_ms    2021-01-11T21:47:08.467980Z # [Note] InnoDB: Using ldnow native AIO
mike_ms    2021-01-11T21:47:08.467992Z # [Note] InnoDB: Number of pools: 1
mike_ms    2021-01-11T21:47:08.467994Z # [Note] InnoDB: Using CPU CRC32 instructions
mike_ms    2021-01-11T21:47:08.468070Z # [Note] InnoDB: Initializing buffer pool, total size = 120M, instances = 1, chunk size = 120M
mike_ms    2021-01-11T21:47:08.470722Z # [Note] InnoDB: Completed initialization of buffer pool
mike_ms    2021-01-11T21:47:08.470680Z # [Note] InnoDB: If the mysqld execution user is authorized, page cleaner thread priority can be changed. See the man page of setpriority().
mike_ms    2021-01-11T21:47:08.470682Z # [Note] InnoDB: Highest supported file format is Barracuda.
mike_ms    2021-01-11T21:47:08.470697Z # [Note] InnoDB: Creating shared tablespace for temporary tables
mike_ms    2021-01-11T21:47:08.470698Z # [Note] InnoDB: Setting file './ibtmp1' size to 12 MB. Physically writing the file full; Please wait ...
mike_ms    2021-01-11T21:47:08.470700Z # [Note] InnoDB: file './ibtmp1' size is now 12 MB
mike_ms    2021-01-11T21:47:08.473128Z # [Note] InnoDB: 96 redo rollback segment(s) found. 96 redo rollback segment(s) are active.
mike_ms    2021-01-11T21:47:08.473132Z # [Note] InnoDB: 32 mergepass rollback segment(s) are active.
mike_ms    2021-01-11T21:47:08.476262Z # [Note] InnoDB: 5.7.16 started; log sequence number 12114362
mike_ms    2021-01-11T21:47:08.476922Z # [Note] Plugin 'FEDERATED' is disabled.
mike_ms    2021-01-11T21:47:08.476962Z # [Note] InnoDB: Loading buffer pool(s) from /var/lib/mysql/ib_buffer_pool
mike_ms    2021-01-11T21:47:08.476978Z # [Warning] Failed to set up SSL because of the following SSL library error: SSL context is not usable without certificate and private key
mike_ms    2021-01-11T21:47:08.476978Z # [Note] Server hostname (bind-address): '*'; port: 3306
mike_ms    2021-01-11T21:47:08.476982Z # [Note] IPv6 is available.
mike_ms    2021-01-11T21:47:08.476987Z # [Note] Server socket created on IP: '*'
mike_ms    2021-01-11T21:47:08.476992Z # [Note] Buffer pool(s) load completed at 210111 21:47:08
mike_ms    2021-01-11T21:47:08.476994Z # [Warning] 'db' entry 'sys mysql.sys@localhost' ignored in --skip-name-resolve mode.
mike_ms    2021-01-11T21:47:08.476996Z # [Warning] 'sys.sys@' entry 'sys@localhost' ignored in --skip-name-resolve mode.
mike_jn    [I 21:47:08.789 NotebookApp] JupyterLab extension loaded from /opt/conda/lib/python3.8/site-packages/jupyterlab
mike_jn    [I 21:47:08.789 NotebookApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
mike_jn    [I 21:47:08.789 NotebookApp] Serving notebooks from local directory: /home/jovyan
mike_jn    [I 21:47:08.789 NotebookApp] Jupyter Notebook 6.1.6 is running at:
mike_jn    [I 21:47:08.789 NotebookApp] http://mike_jn:1234/?token=2fcaae2217e66551c96dc97f39167f4e36586dbd8117
mike_jn    [I 21:47:08.789 NotebookApp] or http://127.0.0.1:1234/?token=2fcaae2217e66551c96dc97f39167f4e36586dbd8117
mike_jn    [I 21:47:08.789 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
mike_jn
mike_jn    To access the notebook, open this file in a browser:
mike_jn    file:///home/jovyan/.local/share/jupyter/runtime/notebook-7-oper.html
mike_jn    Or copy and paste one of these URLs:
mike_jn    http://mike_jn:1234/?token=2fcaae2217e66551c96dc97f39167f4e36586dbd8117
mike_jn    or http://127.0.0.1:1234/?token=2fcaae2217e66551c96dc97f39167f4e36586dbd8117
mike_ms    2021-01-11T21:47:08.478362Z # [Warning] 'tables_priv' entry 'sys_config mysql.sys@localhost' ignored in --skip-name-resolve mode.
mike_ms    2021-01-11T21:47:08.486963Z # [Note] Event Scheduler: loaded 0 events
mike_ms    2021-01-11T21:47:08.487208Z # [Note] mysqld: ready for connections.
mike_ms    Version: '5.7.16' socket: /var/run/mysql/mysql.sock port: 3306 MySQL Community Server (GPL)
```

Use (Changed) Images Together

Worth noting:

- docker-compose file name: „docker-compose.yml“ (*.yaml)
- docker-compose up / down: issued in *.yaml file directory
- Services may error out when started at once
 - start one service at a time:

```
docker-compose up mysql
```

```
docker-compose up jupyter
```

- `Docker-compose down` removes all containers
- Stop container(s) with
`docker container mike_jn stop`

Use (Changed) Images Together

Working with services

- Name container
 - `docker run -p 1234:8888 -name jupyter_ds --rm jupyter/datascience-notebook`
- Use data files (*.csv, etc.)
 - `docker cp some.csv jupyter_ds:/home/jovyan/`
- Persist *.ipynb, etc. outside container
 - `docker run -p 1234:8888 -v C:\Users\Mike_64\Documents:/home/jovyan --rm --name jupyter_ds jupyter/datascience-notebook`