Docker Handout

DSE 512

2022-2-3

Setup

1. Set up your EC2 box
2. Connect (ssh) and update
   ```
   sudo apt update && sudo apt upgrade -y
   ```
3. Install docker
   ```
   sudo apt install -y docker.io
   ```
4. (Optional but handy) Add yourself to the docker group
   ```
   sudo usermod -aG docker $USER
   ```
5. Reboot
   ```
   sudo shutdown -r now
   ```

Running Docker Containers

For some of the examples you will need to run an appropriate ssh tunnel, e.g.

```
ssh aws -L 8888:localhost:8888 -N
```

1. Python
   1. Run
      ```
      docker pull python
      docker run -i -t python
      ```
   2. Python 3 from the container should be loaded. You can `import os` or whatever.
2. R
   1. Run
      ```
      docker pull rocker/r-base
      docker run -i -t rocker/r-base
      ```
   2. R from the container should be loaded. You can run `example(lm)` or whatever.
3. Jupyter (web service)
   1. Run
      ```
      docker pull jupyter/datascience-notebook
      docker run -p 8888:8888 -i -t jupyter/datascience-notebook
      ```
   2. Create the appropriate tunnel
   3. Check the log for a url containing `127.0.0.1:8888`, and paste that into your local web browser
4. Rstudio Server (web service)
   1. Run

```
docker pull rocker/rstudio
docker run -p 8787:8787 -i -t -e DISABLE_AUTH=true rocker/rstudio
```

2. Create the appropriate tunnel
3. Go to http://localhost:8787/ in your local web browser

Creating Docker Containers

Basic steps

1. Create directory
2. Create Dockerfile (text file with some special syntax)

We'll create a small "data science" container with a handful of R and python utilities. We'll start with this as the base of the Dockerfile

```
FROM ubuntu:20.04

ARG DEBIAN_FRONTEND=noninteractive
ENV TZ="America/New_York"

RUN apt-get update && apt-get install -y \
  r-base-dev libopenblas-dev \
  python3 python3-pip python-is-python3

RUN pip install numpy

RUN Rscript -e "install.packages('memuse')"

CMD ["/bin/bash"]
```

Build and run the container via:

```
sudo docker build -t dse512 .
sudo docker run -i -t dse512
```

1. Make sure that you can build the docker image
2. Modify it to install scikit-learn
3. (Quotes) If I modify the `Rscript` line to this, will it still work?

   ```
   RUN Rscript -e 'install.packages("memuse")'
   ```

   Is there any difference between the two? What about this

   ```
   RUN Rscript -e "install.packages("memuse")"
   ```

4. (Dealing with systems dependencies) Using `install.packages()`, install the `openssl` package. It won't work out of the box, but you can find out why and fix it by reading the output.
5. (More complicated dependencies) Using `install.packages()`, install the `tidyverse` package.
6. Create a script in the same path as your Dockerfile called `memuse.r`. Its contents should be

   ```
   memuse::Sys.meminfo()
   ```

   Change the command at the end of the Dockerfile from one that runs bash to

   ```
   CMD ["Rscript", "/memuse.r"]
   ```

ADD the script to your Dockerfile (root path, or modify the `CMD` line appropriately). If you do it right, the script will automatically run when you start the container.

7. It is possible to share files between the host and the container. Read about both:
   - bind mounts https://docs.docker.com/storage/bind-mounts/
   - volumes https://docs.docker.com/storage/volumes/