

Assignment 2

DSE 512

Assigned 2022-2-15 — Due 2022-2-26 11:59pm

Your goal will be to create a Singularity container that you upload to a shared space on ISAAC and use to run a small job on the cluster. You can build the container wherever you like (I recommend an EC2 instance that you terminate afterwards). And you can use whatever you like for software management within the container (e.g. `apt-get`, `pip`, `conda`, install from source, ...), *so long as* all other requirements are satisfied and no extra steps are required to run software in your container.

1. (40%) Create a Dockerfile that uses `ubuntu:20.04` as the base image and installs:

- R (4.x)
- R packages: `tidyverse` and `memuse` R packages (both are on CRAN)
- Python (3.x)
- Python packages: `numpy`, `pandas`, `scikit-learn`

List the contents of your Dockerfile.

Hints: If using `apt`, I recommend this strategy for installing R and the R packages <https://cran.r-project.org/bin/linux/ubuntu/>. If this seems too easy and you want a challenge, try installing everything from source. If using `conda`, make sure you set it up so that if one boots the container to a `bash` shell that entering `R` or `python` launch the required program without needing any additional steps. You can check the R and Python versions by using the flag `--version`, e.g. `R --version`.

2. (10%) Build the Docker image. Call it `dse512`. List the commands you run to build this image.
3. (10%) Convert the Docker image to a Singularity image. Name this image `dse512_${NETID}.sing`. So for example, mine would be `dse512_mschmid3.sing`. List the commands you run to build this image.

Hint: Just rename the image after you build it.

4. (10%) Create a directory in the course project space in lustre on ISAAC (project number UTK0188). The directory should be named your UT netid. List the path.
5. (10%) Upload your singularity image to this directory. List the command(s) you run to transfer the file.
6. (10%) Create a job script in your project directory that will run an R “hello world” (`print("hello world")` will do) using your singularity image. The job should run from within your directory in the lustre project directory. If you can not use the project account UTK0188, then use the opportunistic account UTK0011.
7. (10%) Run the job. The output should now be in your directory.

Because everything is structured incrementally, if for some reason you can't complete an item, you are free to use a proxy. For example, if you get stuck at item 1, then you can grab any container you like as a proxy to continue on.

Everything you need has been covered in class (mostly in the slides and the Docker handout). But you are free to use the internet so long as you don't post any questions yourself (anything already asked and answered is fine). You can also consult with the instructor, the ISAAC support staff (for ISAAC issues), and you can offer limited help to each other (small questions are fine; copying files is not).