THE UNIVERSITY OF CHICAGO | Research Computing Center

Computational reproducibility through virtual environment and Containerization

Parmanand Sinha

Feb 16, 2023

Materials available at https://github.com/rcc-uchicago/singularity-demo

WHERE TO GET THIS PRESENTATION

1. Web browser:

   `https://github.com/rcc-uchicago/singularity-demo`

2. GIT:

   ```
   git clone https://github.com/rcc-uchicago/singularity-demo.git
   ```

LEARNING OBJECTIVES

1. Version Control Systems
2. Use Conda to create environments and install software on the Midway HPC system
3. Learn how to use containers for code portability and reproducibility
4. Some common containerization tools
5. How to find and run containers built by other people.
6. How to build their own container.
7. How to distribute their container online.

URL: https://github.com/rcc-uchicago/singularity-demo

WHAT IS SINGULARITY/APPTAINER

- Singularity is a container specifically designed for HPC
- A container allows to run inside a Linux host a separate instance of Linux that can be completely different distribution/version and have completely different software/environment inside. When a program runs inside a container, it does not depend on host OS or software.
- Container vs Virtual Machine:
  - Container shares kernel with the host, VM does not
  - As a result, container has direct access to the same hardware devices as the host: therefore no problem using GPU, infiniband, GPFS, etc.
  - VM abstracts the hardware, introduces overhead, makes some hardware unavailable inside VM
  - Container is much more lightweight than VM, almost native

WHAT IS SINGULARITY

- Singularity vs Docker
  - Singularity was designed for HPC by the same guy, Gregory Kurtzer from LBNL, who also started CentOS Linux distribution
  - Docker was designed for enterprise, to run trusted applications by trusted users, not suitable for HPC where 1000s users are doing crazy things on the same system
  - Docker requires running a daemon, Singularity does not
  - Docker allows a user to become root and therefore is not secure to run on the system with multiple untrusted users
  - Singularity does not escalate priviliges: inside container you are still the same user as outside, with the same privileges.
- Singularity can use Docker containers either directly or by converting them into Singularity images

WHEN TO USE SINGULARITY

- We started using Singularity couple years ago when users were asking us to install TensorFlow which requires newer version of glibc than the one that midway 1 has. To upgrade glibc one needs to upgrade the whole Linux distribution. This is not so simple to do on such a big system and is not done often. Most of the HPC centers I know are still running version 6 of RedHat/CentOS/Scientific Linux. Singularity allowed us to run TensorFlow inside Scientific Linux 7 container on Scientific Linux 6 host
- Use case #1: to install software that would not run on the version of Linux you have to use but can run on a different version
- Some software stacks are quite complicated to build from scratch on your own and the authors might distribute them as Docker/Singularity images.

WHEN TO USE SINGULARITY

- Suppose you have written a paper and want to preserve or even distribute the exact software environment used to obtain the results that you can still run 10 years from now. Package it into Singularity container.
- Use case #4: to preserve the software environment under which scientific results were obtained.
- Suppose you have rather complicated software that is difficult to build, but you want to run it on different clusters that have different versions of Linux, different software installed, etc. Using Singularity container makes it easy to move between clusters and clouds without having to rebuild everything from scratch.
- Use case #5: to easily run your program on different clusters, clouds without having to recompile

WHERE SINGULARITY IS USED

- Singularity web site has a registry where people who use it can add their site:
  - National labs and computing centers: LBNL, NIH, SDSC, TACC, Army Research Laboratory, ORNL, XSEDE, National Renewable Energy Lab, Centrale Nantes SuperComputing (France) ...
  - Universities: MIT, Caltech, Stanford, Yale, Holland Computing Center at the University of Nebraska, ...
  - Industry: Penguin Computing On Demand, Microway, Genentech, ...
- 2 years after it was created, Singularity became the main container for HPC applications

OTHER WAYS TO BUILD CONTAINERS: VIRTUALBOX

- If you do not run Linux on your desktop/laptop (the best choice!), you can run it as a virtual machine using, for example, VirtualBox
- VirtualBox is available for Windows, Mac, Linux
- Inside VirtualBox you can install any distribution of Linux and have a root access to it
- https://singularity-tutorial.github.io/01-installation/

OTHER WAYS TO BUILD CONTAINERS: WSL

- If you do not run Linux on your desktop/laptop (the best choice!), you can run it as a virtual machine using, for example, VirtualBox
- dism.exe /online /enable-feature /featurename:VirtualMachinePlatform /all /norestart
- dism.exe /online /enable-feature /featurename:Microsoft-Windows-Subsystem-Linux /all /norestart
- wsl –set-default-version 2
- https://singularity-tutorial.github.io/01-installation/

OTHER WAYS TO BUILD CONTAINERS: SINGULARITY HUB

- No need to have a root access or install anything on your computer
- Link your github account with Singularity Hub
- Create a repo, named, for example, FOO and put somewhere under the repo folder `Singularity` or `Singularity.tag` recipe file.
- Link the repo with Singularity Hub.
- After a while your container will be built and become available as `shub://<github user id>/FOO:latest` or `shub://<github user id>/FOO:tag`
- Potential problem:
  - It is very convenient for debugging purposes to be able to write into the image before putting everything into a recipe file. It

XSEDE COMMERCIAL

- eXtreme Science and Engineering Discovery Environment
- https://www.xsede.org
- $\sim$ $120-million project is supported by the National Science Foundation
- $\sim$ 12 clusters in the US
  - Different kinds of clusters: general purpose, GPU, KNL, Hadoop, cloud-like, etc.
- XSEDE allocation is free but you have to submit a proposal like for grant
- Besides research allocation, one can get educational allocation to teach HPC-related class (this tutorial, for example)
- XSEDE provides free training in all aspects of HPC, regardless if you have allocation or not
- XSEDE has representatives on campuses that are called XSEDE

REFERENCES

http://singularity.lbl.gov - Singularity home page

https://singularity-hub.org/ - Singularity Hub

https://hub.docker.com/ - Docker Hub

https://github.com/singularityware/\
singularity/tree/master/examples - Examples of recipes

https://www.virtualbox.org - VirtualBox home page

https://github.com/singularityhub/\
singularityhub.github.io/wiki
- Build containers in the Hub