Using Singularity in HPC environments

On github: https://github.com/stef-mueller/singularity_intro

Goal

- Introducing container as a solution for reproducible software environments
- If anyone is interested to get into touch: I would be happy to work together in building some shared software container

Problem

- Installing complex software environments with numerous dependencies
- Redo for for every user and every HPC cluster with very different problems to solve
- even after successful installation software might break when basic unix libraries are being changed on server

Example

- Example: MetaXcan, successor of PrediXcan

18 lines (17 sloc) | 277 Bytes

```yaml
name: imlabtools
channels:
  - defaults
  - conda-forge
  - moble
  - bioconda
dependencies:
  - python=3.7
  - pandas=0.25.3
  - scipy=1.4.1
  - numpy=1.18.1
  - bgen_reader=3.0.2
  - cyvcf2=0.20.0
  - pyliftover=0.4
  - statsmodels=0.11.1
  - h5py=2.10.0
  - pyarrow=0.11.0

```

https://github.com/hakyimlab/MetaXcan/blob/master/software/conda_env.yaml

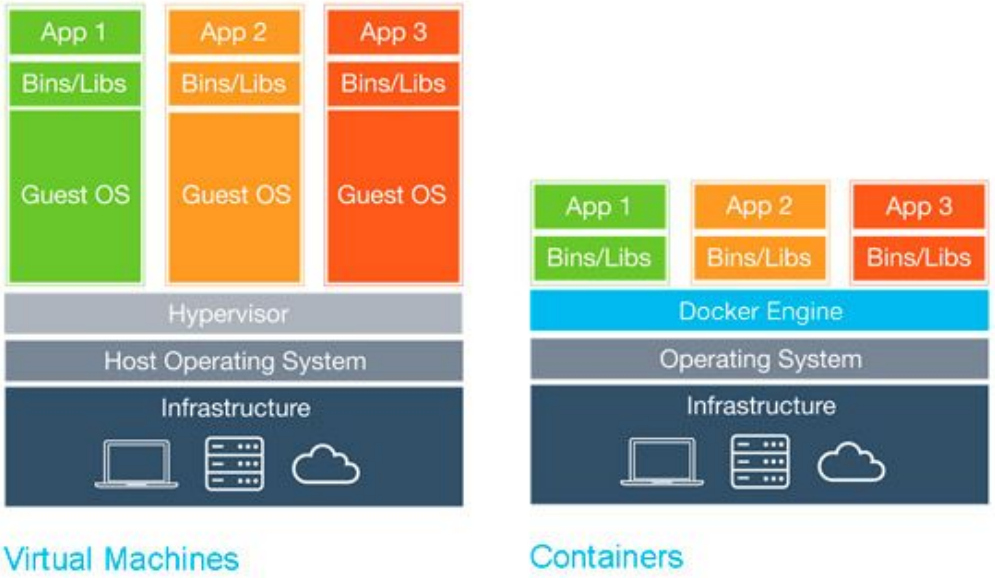
Possible Solutions

1. Share successful installations in shared spaces on individual HPCs
2. Use virtual environments for installation usually using package/env manager conda
3. Containerization: singularity or docker

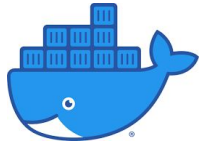
HPC's that support singularity:

- UCL myriad
- UCL CS
- GeL HPC and local
- Most cloud based HPC services (eg. AWS, google cloud)
- also all of your local machines once singularity is installed

What are containers?



- containers use shared operating systems: much more efficient than VMs
- instead of virtualizing hardware like VMs, containers rest on top of a Linux instance

Singularity  vs. docker 

- More commonly used in HPC architecture
- Smaller user base, but can use docker container
- More secure, all processes are executed as the user and with preset user permissions

Service to find/share container

https://singularityhub.com/

- More commonly used in general
- Bigger user base - much more available containers

Service to find/share container

https://hub.docker.com/

Singularity Example

Example1: Performing simple plink2 analysis on CS cluster

Example2 in git repo: perform sample analysis for `SPrediXcan`

Caveats

- container take up more space than simple system install
  - How to change tmp folders for image building when error running out of space:
  - https://ubccr.freshdesk.com/support/solutions/articles/13000065620-singularity-build-error-no-space-left-on-device
- don't blindy trust all containers you might encounter
  - stick ro reputable sources or have look at build file
- creating containers can be a bit tricky but there is lots of sources to find containers
- where to find containers:
  - On github:
    - https://github.com/BioContainers/containers; huge repo of more than 100 bioinfo software container
    - https://github.com/RTIInternational/biocloud_docker_tools
  - On https://hub.docker.com/
  - Or just google: "<software> docker" or "<software> singularity"
  - Chances are somebody already created a container which can be repurposed