

R Users Group

Intro to Docker - Why and How to Use It

David Thole

University of Iowa, College of Pharmacy

2019/01/24



Agenda

1. What is it, how to start using it?

2. How it works

3. Basic Usage

Closing



Agenda

1. What is it, how to start using it?

2. How it works

3. Basic Usage

Closing



Purpose of Docker

Came about from the concept of "Dependency Hell", and "It works on my machine" which is the tightly coupled dependencies that are difficult to reproduce on other machines.



Why use Docker? (System Admin Perspective)

- ① Ability to run PowerShell/Bash on any OS.
- ② Ability to publish workflows that include the dependencies for that specific process, or set of processes.
- ③ Ability to reproduce an environment in a deterministic manner (oops, forgot to install X)



Why use Docker? (Developer Perspective)

- ① Ability to bundle dependencies for a project into one bundle (No more needing version .NET 4.5 runtime installed on system X)
- ② Ability to have containers more accurate between test and production (same container can be used with different entrypoints)
- ③ Ability to abstract dependencies of a development machine (No more install X, then Y, then Z - for each developer, spending a day setting up a new machine)
- ④ Ability to test new software/products quickly
- ⑤ Ability to script out deployment/testing much easier than current solutions



Why use Docker? (DBA/Analyst Perspective)

- ① Ability to test data sets easier (script the loading from a backup)
- ② Ability to create working sets (e.g. working on project X's data, now project Y)



Why use Docker? (Researcher Perspective)

- ① Ability to reproduce results using an exact setup (vs it ran on machine X 3 years ago, and has since been reformatted)
- ② Ability to script an experiment's processing, to enforce standardization, but to aid in reproducibility



Agenda

1. What is it, how to start using it?

2. How it works

3. Basic Usage

Closing



Execution Considerations



Agenda

1. What is it, how to start using it?

2. How it works

3. Basic Usage

Closing



Pulling Containers



Execution Containers



Building Containers



Using Multiple Containers



Agenda

1. What is it, how to start using it?
2. How it works
3. Basic Usage

Closing



4. In Closing

Resources:

- Pluralsight (Docker Path)
- Docker Up & Running (O'Reilly)
- Dockerhub (best place to look at what's out there)

Useful/Fun Containers:

- thedarktrumpet/rstudiojava - Studio, Shiny (For R Use)
- microsoft/mssql-server-linux - Microsoft SQL Server
- jupyter/all-spark-notebook

