The Coffee and Code Workshop Platform, an Integrated System for Analytic Tool Training Workshops and Tool Use

Karl Benedict (kbene@unm.edu) & Jonathan Wheeler (jwheel01@unm.edu)
University of New Mexico, University Libraries, Research Data Services

IN41E-0881
AGU 100 FALL MEETING
Washington, D.C. | 10–14 Dec 2018

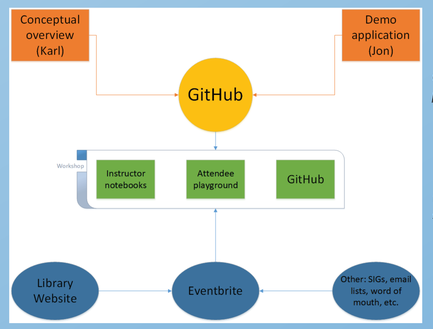
The Problem

Training and skill development in core data management concepts, methods, and technologies include short single-session group workshops, self-paced online learning objects, and multi-day and longer-form training and education programs. ***These training sessions often require the installation and use of specific software on participants computers and the sharing of instructional materials.*** This common need to deliver a complex combination of training materials, sample code, and a software environment within which students can work poses a challenge to training providers and often results in valuable instruction time being spent on training-specific logistics instead of the intended training content. To meet the joint requirements of providing instructional materials in a useful form, providing functional examples, and enabling the streamlined installation and use of the tools covered, UNM's Research Data Services (RDS) program has developed an instruction platform that allows workshop participants to experiment with the tools being presented, and provides them with the ability to continue to run the platform on their own computer for continued experimentation and learning.

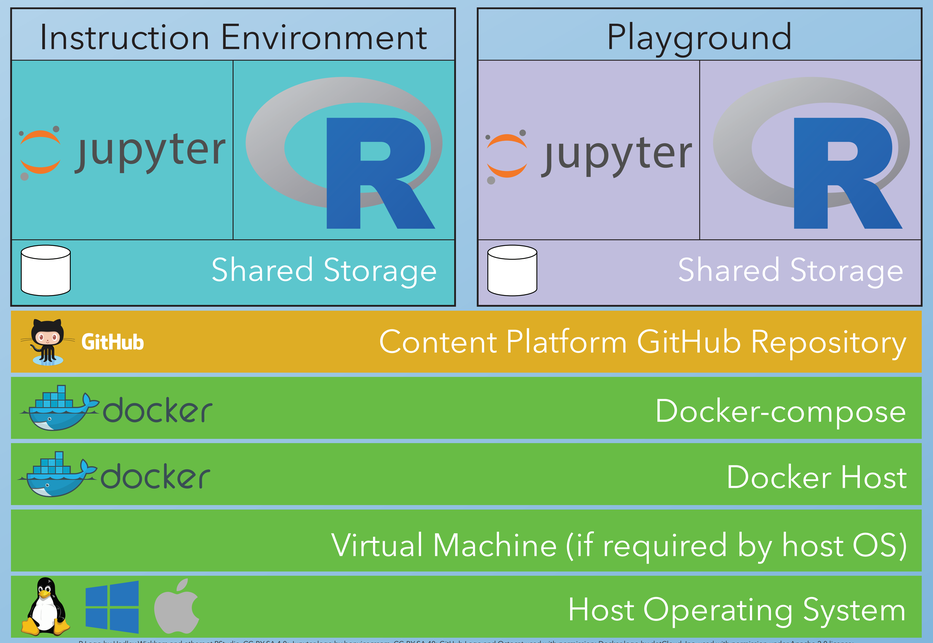
The Context

The Research Data Services team in the UNM's University Libraries has developed a growing collection of training workshops on technical topics including:

- ☑Programming concepts using python
- ☑Version control with GitHub & GitLab
- ☑Document production with Markdown and Pandoc
- ☑Database design basics
- ☑Jupyter Notebooks
- ☑R and RStudio
- ☑Containerization using Docker
- ☑Working from the command line
- ☑... and others



Common workshop development, promotion, and presentation components. Developed content is managed within GitHub as a shared development environment; links to workshop materials in GitHub are shared as part of the workshop promotion process; separate copies and platforms are used for the workshop when presented.



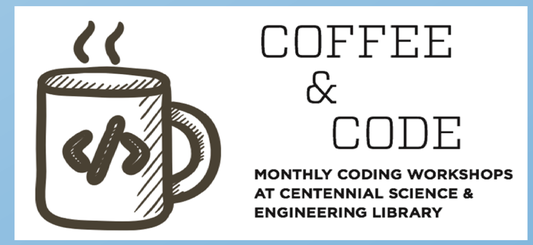
R Logo by Hadley Wickham and others at RStudio, CC-BY-SA 4.0; Jupyter logo by bonyicecream, CC-BY-SA 40; GitHub Logo and Octocat used with permission; Docker logo by dotCloud, Inc. used with permission under Apache 2.0 license; Tux loco by Larry Ewing, Simon Budig, Garrett LeSage, used with permission

Platform components when running on a remote or local system. From the bottom of the diagram: (1) the operating system of the computer running the platform, (2) a virtual machine running on the host platform if the host OS does not have native support for the Docker Host environment, (3) the Docker Host environment within which individual Docker containers are run, (4) the additional Docker-compose tool that is an additional installation on some system depending on the default Docker installation process, (5) the content of the platform GitHub repository that contains all of the needed configuration and execution files to run the platform, and (6) the paired instruction and playground environments that can be run to allow for a separation between the content of the instruction system and the system within which workshop participants are experimenting.

Platform Capabilities

- ☑Streamlined deployment onto any platform that can support the Docker system
- ☑Automatic retrieval of all workshop materials (customizable and based on content being separately managed in GitHub or other online system) during platform start
- ☑Capability of updating/replacing all current workshop materials in the environment while running - streamlined development to operations iteration
- ☑Separate environments for instructors and workshop participants reducing risk of instruction content being impacted by participant actions
- ☑Customization of environment to support additional tools that can run in a Linux environment - current support for Jupyter Notebook, RStudio, Pandoc & markdown

UNM Research Data Services GitHub Repository Collection
https://github.com/unmrds

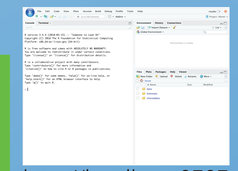


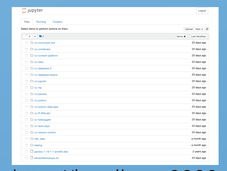
Deployment Workflow

Download & Install Docker on Host Computer — https://dockr.ly/2PeNuu3

Download Zip File or Clone Repository to Host Computer — https://github.com/unmrds/cc-content-platform

Run `docker-compose` *Command on Host*


http://localhost:8787


http://localhost:8888

hosted playground instances

Acknowledgements

The development of the platform and associated workshop content has been partially funded by UNM's University Libraries; NSF Awards 1757207, 1301346, 1329470; and IMLS Grant LG-70-18-0092-18.