Name:	Lee Evans, Marc Suchard, Jon Duke
Affiliation:	LTS Computing LLC, University of California, Regenstrief Institute
Email:	levans@ltscomputingllc.com
Presentation type (select one):	Poster

Broadsea – The OHDSI Open Source Standard Software Stack Packaged as Docker Container Images for Cross-Platform Installation

Lee D. Evans[1], Marc A. Suchard, MD, PhD[2], Jon D. Duke, MD, MS[3]
[1]LTS Computing LLC, West Chester, PA; [2]Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA; [3]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN;

Abstract

We packaged the OHDSI open source standard software stack into Docker containers with the aim of simplifying cross-platform OHDSI software installation on a range of Operating Systems, DBMSs and infrastructure. We believe this simpler software deployment option will help encourage OHDSI Community members to download and install the full OHDSI software stack for research on their own CDM databases and more easily participate in OHDSI Network studies.

The software container image build process and the OHDSI container images are collectively known as Broadsea. The OHDSI Broadsea Docker containers may be configured to connect to an OMOP Common Data Model Version 5 database.

Introduction

The OHDSI software stack deployment process requires some familiarity with the underlying Java/HTML5/R development technologies. Further, a Tomcat application server and an R environment must be available along with the CDM database. Deployment is a typical Java/HTML5/R application process but there are a number of dependencies and a number of steps that need to be performed.

We use continuous integration and Docker container technology to build and package the open source OHDSI web applications and OHDSI R packages into cross-platform software containers. The Tomcat application server, R/RStudio environment, OHDSI web applications and OHDSI R packages are already included in the Broadsea containers, simplifying the installation process.

The Broadsea Docker containers are configured by editing the docker-compose.yml file to connect to the local Common Data Model Version 5 database, and the database sources and source daimon tables are populated by editing and executing a SQL script file.

The Broadsea containers have a dependency that the Docker engine and Docker compose utility (installed together using docker-toolbox on Windows/Mac and separately on Linux) must be installed on the host machine.

Docker Container Technology Background

A software application is packaged along with run-time dependencies into a single portable binary file called a Docker Image. The Docker engine service running on the host machine will execute the Docker image using a lightweight shared kernel approach. An instantiated Docker image is known as a Docker container. A Docker container has a life cycle where it is run; the container may be stopped and re-started, and eventually it is removed.

A "Dockerfile" file is used to define and build a Docker container image. A "docker-compose.yml" file is used to set the configuration parameters and orchestrate the runtime behavior of one or more Docker containers. We have developed the Dockerfiles and docker-compose.yml files to build and run the OHDSI Broadsea containers.

The OHDSI Broadsea Docker Images

We have produced two Docker images and more may be added in the future as new OHDSI tools are developed:

- OHDSI Web Applications (container image developed by LTS Computing LLC)
    - Tomcat web application server with the OHDSI Web API installed
    - Atlas
    - Calypso
    - Penelope (still under development)
    - Achilles generated reports for the SynPUF 1k simulated patient sample dataset (data reports visible in Atlas)
- OHDSI R (container image developed by Marc Suchard based on the Rocker RStudio Docker image)
    - RStudio Server web application with the OHDSI R packages installed

Operating Systems & Databases Supported

The OHDSI Broadsea Docker containers will run on any OS that supports Docker containers, including Linux, Windows and Mac OS X. The containers may be used with any DBMS that is supported by the OHDSI SqlRender technology including SQL Server, Oracle and PostgreSQL. Proprietary jdbc drivers are not included in the Docker containers but they will be used if they are made available in the directory where the containers are launched.

Conclusion

The goal of Broadsea is to simplify the installation of the full OHDSI standard software stack for the OHDSI Community.

Broadsea provides a standard release distribution of the OHDSI software stack that should facilitate future OHDSI research reproducibility[1]. A researcher wanting to reproduce an OHDSI study with their own CDM data can use the same release of the OHDSI software stack that was used to run the original study if it was preserved as a Broadsea Docker image.

There are also future opportunities to use additional OHDSI Broadsea Docker containers to run additional OHDSI tools and backend services (e.g. R Service Bus / RPooli) and to take advantage of performance scaling using Docker engines on cluster computing platforms such as Mesos, Kubernetes and Docker Swarm, running on local infrastructure or in the cloud (e.g. Amazon Web Services or Google Cloud Platform).

References

1. Boettiger C, An introduction to Docker for reproducible research. CoRR. 2014;abs/1410.0846