DataAPI 01A Docker SDB on Linux
Tod Casasent
2020-06-23-1200

Introduction

These instructions are aimed at people familiar with Linux administration and GitHub, Docker and Docker-Compose, with root access to their Linux environment. Installation is for external users.

The Docker SDB is an image for viewing, searching, and downloading Standardized Data archives. Standardized Data archives are ZIP files containing batches.tsv and clinical.tsv files, a matrix_data.tsv file, and an index.json file used by the Standardized Data Browser (SDB).

Target Operating System and Installation

These instructions were tested on RHEL 7. These instructions with appropriate modifications should work as a basis for installing on other distributions. Perform the installs in the order given in this document. Some steps are dependent on previous steps. All elements of this document expect a Linux host. It may be possible to convert these instructions to Windows or OSX, but no such ability is supported.

Operating System Prerequisites

Docker and Docker-Compose are required installs.

$ docker --version

Docker version 19.03.11, build 42e35e61f3

$ docker-compose --version

docker-compose version 1.25.4, build 8d51620a

Clone the GitHub Repository

Clone the GitHub repository with a shallow clone, since you will not be checking anything back in. This clone call grabs the newest version from master.

```
git clone --depth 1 https://github.com/MD-Anderson-Bioinformatics/DataAPI.git
```

This places the cloned code into a new DataAPI directory.

5.

File Preparation from Clone

1.

Optional: Compile

Compiling will be documented in a future release. The apps/DAPI Netbeans 11 project is used for this application.

Required: Copy to Base Directory

Since Docker Containers do not save internal modifications between runs, SDB expects the following directories:

/DAPI/DATA /DAPI/CONFIG /DAPI/INDEXES /DAPI/LOGS

The contents of DATA will be explained in the future, but contain the ZIP archives referred to in the INDEXES. LOGS stores the log files from Tomcat.

CONFIG contains two files. Both are optional, but when used with the Batch Effects Viewer, the lack of dapi.properties will significantly affect usability.

The dapi.properties file contains XML with URLs for the three possible applications used together with the DAPI application.

<?xml version="1.0" encoding="UTF-8" standalone="no"?>

<!DOCTYPE properties SYSTEM "http://java.sun.com/dtd/properties.dtd">

<properties>

<entry key="BEV_URL">http://localhost/BatchEffectsViewer</entry>

<entry key="BQF_URL">http://localhost/BEVQF</entry>

<entry key="STD_URL">http://localhost/StandardizedDataBrowser</entry>

</properties>

The dapi-filter.tsv contains tab-delimited data that allows remapping of GDC names into more use-friendly, coordinated terminology.

Derivations Category Platform Details Category-New Platform-New Details-New

current Copy Number Segment DNAcopy Copy Number DNAcopy With CNV

current Masked Copy Number Segment DNAcopy Copy Number DNAcopy No CNV

legacy Copy number variation Affymetrix SNP Array 6.0-hg19-nocnv Copy Number SNP6

current Methylation Beta Value Liftover noXY DNA Methylation Combined Methylation No Sex Chromosomes

current Methylation Beta Value Liftover wXY DNA Methylation Combined Methylation With Sex Chromosomes

legacy DNA methylation Illumina Human Methylation 27 noXY DNA Methylation Methylation 27 No Sex Chromosomes

legacy DNA methylation Illumina Human Methylation 27 wXY DNA Methylation Methylation 27 With Sex Chromosomes

legacy DNA methylation Illumina Human Methylation 450 noXY DNA Methylation Methylation 450 No Sex Chromosomes

legacy DNA methylation Illumina Human Methylation 450 wXY DNA Methylation Methylation 450 With Sex Chromosomes

current Gene Expression Quantification HTSeq - Counts Gene Expression HTSeq - Counts

current Gene Expression Quantification HTSeq - FPKM Gene Expression HTSeq - FPKM

current Gene Expression Quantification HTSeq - FPKM-UQ Gene Expression HTSeq - FPKM-UQ

legacy Gene expression RNA-Seq-gene-unnormalized-v2 Gene Expression RNA-Seq-gene-unnormalized-v2

legacy Gene expression RNA-Seq-isoform-unnormalized-v2 Gene Expression RNA-Seq-isoform-unnormalized-v2

legacy Gene expression RNA-Seq-v1 Gene Expression RNA-Seq-v1

current Isoform Expression Quantification BCGSC miRNA Profiling miRNA miRNA-Seq Isoform Quantification

current miRNA Expression Quantification BCGSC miRNA Profiling miRNA miRNA-Seq Gene Quantification

legacy miRNA-Seq miRNA gene quantification-gene-hg19-miRNA miRNA miRNA-Seq Gene Quantification

legacy miRNA-Seq miRNA isoform quantification-hg19-isoform-miRNA miRNA miRNA-Seq Isoform Quantification

legacy Simple somatic mutation DNA-Seq-Illumina MiSeq Mutations DNA-Seq-Illumina MiSeq

current Gene Level Copy Number Scores GISTIC - Copy Number Score Mutations GISTIC Copy Number

current Masked Somatic Mutation MuSE Variant Aggregation and Masking Mutations MuSE Somatic

current Mutations Call Analysis Mutations MutBatch Analysis

legacy Mutations Call Analysis Mutations MutBatch Analysis

current Masked Somatic Mutation MuTect2 Variant Aggregation and Masking Mutations MuTect2 Somatic

current Masked Somatic Mutation SomaticSniper Variant Aggregation and Masking Mutations SomaticSniper Somatic

current Masked Somatic Mutation VarScan2 Variant Aggregation and Masking Mutations VarScan2 Somatic

legacy Protein expression MDA_RPPA_Core Protein Expression RPPA

1.

Setup Local Information from Base Directory

1.

Template Dockerfile Settings

The Dockerfile template file should be updated and renamed.

Replace <IMAGE_NAME> with the IMAGE_NAME used below. This is informational on the image only.

Replace <RELEASE_VERSION> with "latest" or some other version number. This is informational on the image only.

Replace <USERID> with a Linux UID that has access to the /DAPI directories described elsewhere.

Template docker-compose.yml Settings

The docker-compose template file looks like this. Is should be updated and renamed.

this is the docker-compose version

version: '3'

file version 2020-06-23-1200

services:

<IMAGE_NAME>_service:

use existing default network

network_mode: bridge

restart this container if it crashes

restart: always

build:

build from directory in context and Dockerfile

context: .

dockerfile: Dockerfile

container_name: <IMAGE_NAME>_cont_<ENVIRON>

update :latest to desired version

image: <IMAGETXT>:<DESIREDTAG>

volumes:

outside access for data files outside:inside

- <INDEXPATH>:/DAPI/INDEXES:ro

- <CONFIGPATH>:/DAPI/CONFIG:ro

- <OUTSIDE_GDC_DATA_PATH>:/DAPI/DATA/GDC:ro

- <LOGPATH>:/opt/tomcat/logs

read-only file to set time and timezone to same in image as on server

- /etc/localtime:/etc/localtime:ro

- /etc/localtime:/etc/timezone:ro

ports:

(outside)host port:container port(inside) for Tomcat

outside/host port is only set here (other docker compose have ports in more than one place)

- "<OUTSIDE_PORT>:8080"

tty: true

Below, find the all capital text (surrounded by pointy-brackets < and >) in the Docker Compose Template and replace with the information described below.

IMAGE_NAME	The container name is described in the file as "<IMAGE_NAME>_cont
ENVIRON	The container name is described in the file as "<IMAGE_NAME>_cont
IMAGETXT	The image used by the Docker Compose file is "<IMAGETXT>:<DESIR
DESIREDTAG	The image used by the Docker Compose file is "<IMAGETXT>:<DESIR
INDEXPATH	This gives the path to the index file. For these directions, use /DAPI/IN
CONFIGPATH	This gives the path to the config files. For these directions, use /DAPI/C
OUTSIDE_GDC_DATA_PATH	This gives the path to the data (ZIP archives). For these directions, use /
LOGPATH	This gives the path to the log directory. For these directions, use /DAPI/
OUTSIDE_PORT	Use the outside port you wish to use for connection to the application. (T

Dockerfile Settings

The Dockerfile uses the installations directory at docker-build/SDB/installations. The apps/DAPI Netbeans 11 should be compiled, and the resulting WAR renamed to StandardizedDataBrowser.war and placed in the installations directory.

Additional build docs will be released in the future.

2.

Images

4.

Recommended: Pull Images from Docker Hub

In the directory with the update docker-compose.yml file using the name given it when editing, pull the image with:

```
docker-compose -f docker-compose.yml pull
```

Starting and Stopping the Docker Compose Stack

In the directory with the docker-compose.yml file, the container is started using:

```
docker-compose -p EXT -f docker-compose.yml up -d
```

The EXT may be varied if needed on your system to ensure unique ids for the container.

The Docker Compose container can be stopped using:

```
docker-compose -p EXT -f docker-compose.yml down
```