BEIStack 01A Installing External on Linux
Tod Casasent
2019-02-05-1200

1 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.

2 Target Operating System and Installation

These instructions were tested on Debian 9.1 installed on Oracle VirtualBox. These instructions with appropriate modifications should work as a basis for installing MBatch 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.

3 Operating System Prerequisites

Docker and Docker-Compose are required installs.

Docker versions used in development and testing were:

Version: 17.12.0-ce or 17.06.0-ce

API version: 1.35 or 1.30

Go version: go1.9.2 or go1.8.3

Docker-Compose versions used in development and testing were:

docker-compose version 1.16.1 or 1.14.0

docker-py version: 2.5.1 or 2.3.0

CPython version: 2.7.13

4 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/BatchEffectsInterfaceStack.git
```

5 File Preparation from Clone

5.1 Optionally Compile

If you want to use pre-built images from Docker Hub, you can skip compiling. Otherwise, run the scripts/00_compile.bash script.

5.2 Copy to Base Directory

Since Docker Containers do not save internal modifications between runs, the Batch Effects Interface Stack expects external locations for storing data. The scripts/01_copy.bash file takes the full path to the desired directory. In this document, we will use the path /BEI_EXT for the base directory. So, run:

```
./01_copy.bash /BEI_EXT
```

This script also creates the directories used by the Docker Compose Stack. Once done, continue the rest of this from the /BEI_EXT directory.

6 Setup Local Information from Base Directory

6.1 Dockerfile and docker-compose.yml Settings

Since Docker Containers do not save internal modifications between runs, the Batch Effects Interface Stack expects external locations for storing data. Other setup information is also needed.

The file /BEI_EXT/build/02_sed.bash has the following variables which need to be edited.

- BASEDIR This is the base directory. For this example, use /BEI_EXT
- DESIREDTAG This is the desired download tag for the images. Use the newest one listed or one with the version you desire.
- SUBNET This is the subnet for the Docker Stack created to run Batch Effects Interface. If you do not need or use subnets, you can leave this as is. If it is not needed, edit /BEI_EXT/build/bei-stack/docker-compose_template.yml and delete the ipam, config, and subnet lines under networks.
- BEI_URL This is the base URL for the Batch Effects Interface HTTP GUI. You do not need to use http:// here.
- BEV_URL This is the base URL for the Batch Effects View which is part of this stack. You do not need to use http:// here.
- SERVER_TITLE This is a substring displayed in the header GUI for the Batch Effects Interface, useful for labelling different environments.

The file /BEI_EXT/build/02_sed.bash has the following variables which can be optionally edited.

- USER_ID If you are building the images yourself and need a different user id to access the external directories (in this case the BEI_EXT sub-directories), edit this entry.
- BEV_PORT If you wish to change the default Batch Effects Viewer port, edit this entry. This does not require a rebuild, since it is set by Docker Compose.
- BEI_PORT If you wish to change the default Batch Effects Interface port, edit this entry. This does not require a rebuild, since it is set by Docker Compose.

In the following examples, I made these changes, in addition to removing the subnet settings from /BEI_EXT/build/bei-stack/docker-compose_template.yml:

```
BASEDIR="/BEI_EXT/"
DESIREDTAG="2019-02-04_BEIStack"
SERVER_TITLE=" External BEI"
…
BEI_URL="localhost:${BEI_PORT}"
BEV_URL="localhost:${BEV_PORT}"
```

The networks section (and first part of services) looks like this:

```
networks:
  # network for this compose stack
  bei_network_<ENVIRON>:
    driver: bridge
services:
  # beiService runs tomcat with Batch Effects Interface and Batch Effects Viewer
  beiService:
```

6.2 Create Dockerfile and docker-compose.yml Files

Run the script /BEI_EXT/build/02_sed.bash.

7 Images

7.1 Optionally Build Images

If you wish to build packages (in which case I recommend you change the DESIREDTAG to avoid confusion), run the script /BEI_EXT/build/03_build.bash.

7.2 Pull Images from Docker Hub

In the /BEI_EXT/build/bei-stack/ directory with the docker-compose.yml file, pull the images with:

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

8 Fine-Tune Settings

Setting for BEI are stored in /BEI_EXT/docker_data/PROPS/bei.properties and /BEI_EXT/docker_data/OUTPUT/gdc.properties.

8.1 Entries for gdc.properties

Since machines have different amounts of memory, we provide a configurable memory setting for downloading GDC data. The MEMSIZE variable should be 64GB to download any currently supported/known data set. (Data sets may increase in size over time.)

Recommended MEMSIZE settings are from 10G to 20G or 64G.

```
# memory size to use for running GDC downloads
# 10-20G works for smaller data sets. Larger datasets may use up to 64G
MEMSIZE=20G
```

8.2 Manual Entries for bei.properties

Similary, the bei.properties file can be manually configured if needed. The provided file uses the settings for Development Values.

Key	Description	Development Value	Production Value
maxDscThreads	The number of threads to be used calculating DSC values. This requires memory and CPU time.	5	10
maxBoxplotGeneCount	The maximum number of genes to use in the boxplot. We strongly recommend a maximum of 5000 for most genomic data sets.	2500	5000
minBoxplotGeneCount	The minimum number of genes to use in the boxplot. We strongly recommend using the maximum value, unless you are simply testing the system.	500	1000
maxDscGeneCount	The number of genes used for calculating the DSC. We strongly recommend using at least a maximum of 10,000 for production systems.	5000	10000
maxDscPermutations	The number of permutations used for calculating the DSC p-value. We strongly recommend using 2000.	2000	2000
filterMaxAllowed	The maximum number of cells (rows times columns) in a matrix used for processing Batch Effects. This is useful for systems limited by memory.	4000000	8000000
filterMinAllowed	The minimum number of cells (rows times columns) in a matrix used for processing Batch Effects. This is useful for systems limited by memory. NOTE: A value of 0 means "use all data".	1000000	0

9 Starting and Stopping the Docker Compose Stack

In the /BEI_EXT/build/bei-stack/ directory with the docker-compose.yml file, the containers (stack) are 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 stack.

The Docker Compose Stack can be stopped using:

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

10 Debugging and File Maintenance

To connect to the different containers as root, using the following:

```
docker exec -t -i -u 0 bei_cont_EXTERNAL /bin/bash
docker exec -t -i -u 0 gdc_cont_EXTERNAL /bin/bash
docker exec -t -i -u 0 bei_cont_EXTERNAL /bin/bash
```

To connect as the default user (1011 docker_tcga), use:

```
docker exec -t -i bei_cont_EXTERNAL /bin/bash
docker exec -t -i gdc_cont_EXTERNAL /bin/bash
docker exec -t -i bei_cont_EXTERNAL /bin/bash
```

The various tomcat logs are in: /opt/tomcat/logs/. A future version will place these logs outside the container.

Another useful command is to use the container root to set access to directories and files created by the container to 777.

```
cd /BEI_EXT
chmod -R 777 *
```

11 Testing the Docker Compose Stack and URL

The URL for your install will be something like:

http://your-server.your-company.com:8181/BatchEffectsInterface/

To test the Docker Compose Stack, see the BEStack 02A Using Batch Effects Interface Assessments document and follow the Downloading GDC Data, Configuring Assessments, and running the job sections.